AD 252 000

EVOLUTION OF THE ASTIA AUTOMATED SEARCH AND RETRIEVAL SYSTEM

ASTIA

JANUARY 1961

ARMED SERVICES TECHNICAL INFORMATION AGENCY

The Scientific Information Center of the Department of Defense
An Activity of the Air Force Air Research & Development Command

EVOLUTION OF THE ASTIA AUTOMATED SEARCH AND RETRIEVAL SYSTEM

JANUARY 1961

AD 252 000

BY
WILLIAM HAMMOND
LT. COLONEL, USAF

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

CONTENTS

FOREWORD

This report, like its predecessors,* has been compiled in response to the widespread interest expressed in ASTIA's automation program, and particularly in its machine search and retrieval system. It differs from the earlier reports, however, in that it concentrates to a great degree upon the machine considerations, as such, encountered in planning, developing, and operating an automated retrieval system.

It is expected that this rather specialized approach will be of greatest value to those concerned with the actual devising and programming of similar or related efforts. But, insofar as the report sheds further light on the intricate relationships between document services and machine system, it is hoped that the presentation will be of substantial interest to all concerned with those relationships.

William Hammond

WILLIAM HAMMOND
Lt. Colonel, USAF

---

* AD-227 000 - "Automation of ASTIA...A Preliminary Report," dated 1 December 1959, and AD-247 000 - "Automation of ASTIA - 1960," dated December 1960.

ABSTRACT

This report covers the evolution of the ASTIA automated search and retrieval system. Early consideration and punch card approaches are treated briefly. The current system and programs are described in detail with the exception of the creation of the Thesaurus which is covered in two other ASTIA publications -- AD-227 000 and AD-247 000. The programs discussed are for the UNIVAC Solid State 90 magnetic tape system. Flow charts are included of the program now in use which executes ten simultaneous searches to four levels of coordination. Modification of this program to provide for either ten six-level coordinations or any combination of a 60-descriptor coordination are also discussed.

EARLY CONSIDERATIONS

Interest in machine compilation of bibliographies from the vast collection of military research and development documentation began shortly after World War II. Prior to 1949, the custodian of the collection, the Air Documents Division,* and its successor organization, the Central Air Documents Office (CADO)* had experimented with mechanization. They tried McBee notched cards, punch cards, addressograph plate registers, and photographic reductions on microfilm. From this experimentation, it was concluded that the employment of any of these media would provide little if any improvement over the manual operation. However, the collection of documents was growing at a rapid rate, and the investigation of mechanization techniques continued.

In January of 1949, consultants of IBM assisted the CADO staff in further experimentation with punch cards to mechanize bibliography service. The major shortcomings of a punch card information search and retrieval system soon became apparent. The manipulation of a great volume of punch cards would be cumbersome, time consuming and more costly than the manual operation it replaced. Further investigation into a punch card retrieval system was abandoned by CADO.

When ASTIA was established in May of 1951, it inherited from CADO the responsibility for providing timely bibliography service

---

* Former USAF organizations.

from a collection of military research and development documentation that was still growing at the rate of 30,000 new titles a year. The director of the new organization recognized this responsibility in his initial plan of operations as follows:

"Mechanical and electronic storage and retrieval of bibliographical information appears to be a requirement for the vast volume of technical information being generated currently. In addition, new concepts of indexing compatible with both machine and manual operations must be perfected.

"Unless new mechanized techniques of storage, indexing and transmission of technical reports are developed, the cost of providing adequate services may soon become prohibitive."

Shortly after its establishment, ASTIA sponsored a contract with Dr. Mortimer Taube of Documentation Incorporated to make a study of existing subject cataloging systems, and evaluate their adequacy for use by ASTIA. This study led to the conclusion that no existing system was really suitable. Further work undertaken as part of the same contract effort resulted in the Uniterm System of Coordinate Indexing.

The uniterm retrieval system was language oriented. It consisted of a collection of words used to characterize the actual content of a document. At the option of the cataloger, new uniterms were added as they appeared in documents being added to the collection. The meanings of the terms conformed to normal usage as defined by dictionaries.

Due to the rapid growth of the AD collection, the free use of uniterms soon became unmanageable from a retrieval standpoint. The need for some degree of control and for a standard term dictionary -- or thesaurus -- was soon recognized.

At first, it was believed that the uniterm system could be operated without the application of machines. The system was also considered to offer great promise in terms of its susceptibility to automation. However, it was soon realized that retrieval of information from a large and diversified collection would require some form of mechanization, if not complete automation. In spite of these shortcomings, the potential of the uniterm system was recognized. In anticipation of a machine system ultimately becoming available, uniterm assignment to all new documents being processed into the ASTIA system became a standard procedure in early 1953. This proved to be a farsighted decision, since it provided a great amount of the raw data for creating the machine retrieval files for the current system.

In the fall of 1954, a contract was let with Airborne Instruments Laboratory to study ASTIA operations and existing data processing equipment to determine whether any equipment was available that could be profitably employed to automate ASTIA's entire operation. The conclusion of this study was that the commercial data processing equipment then available could not be profitably employed in ASTIA operations. The study also found that too little was known at that

point in time about the information retrieval functions to warrant any action to apply data processing equipment to that area. Based on the findings and the conclusions, the contractor recommended that a full scale study be undertaken which would formulate specifications for a special purpose data processing system suitable to ASTIA's needs. This recommendation was not acted on since there appeared to be little likelihood of obtaining the resources to develop a special-purpose data processing system unique to ASTIA's needs.

ASTIA staff personnel continued to keep informed on developments in the data processing field, including the development of character-reading devices, random access memories, and related equipment. The advice and assistance of professional personnel and consultants were solicited on a continuing basis.

PLANNING FOR THE EMPLOYMENT OF THE SPECIAL INDEX ANALYZER PUNCH CARD SYSTEM (IBM 9900)

Shortly after ASTIA had been relocated at Arlington Hall Station in February 1958, a fresh look into automation was undertaken. Systems personnel of the IBM Corporation conducted a survey of ASTIA with the objective of determining whether the state of the art had progressed to the point where any commercially available data processing systems could be profitably utilized.

The IBM survey concluded that the majority of the functions susceptible to automation revolved around request processing and inventory control. These were considered business-type applications that could be handled by commercially available general-purpose data processing systems.

For compiling report bibliographies, the IBM survey recommended the Special Index Analyzer (IBM 9900), an outgrowth of the "COMAC" information retrieval concept generated by Documentation Incorporated and under contract to the Office of Scientific Research, Headquarters USAF.

The concept developed jointly by the IBM data processing consultants and the ASTIA staff would have provided the Washington Regional Office and the four decentralized Regional Offices of ASTIA with a punch card machine search and retrieval capability.

An adaptation of the Uniterm System of Coordinate Indexing was contemplated at each Regional Office, with reserve capability at the central facility to assist in peak periods. One Special Index Analyzer was recommended for use at each Regional Office.

It was planned to mechanically edit the uniterms assigned to the 175,000 documents then in the AD collection. From this edited list of uniterms, a thesaurus-type dictionary would be developed. The acceptable terms in the controlled vocabulary would be mechanically reassigned to each document. The master file of search term cards would then be automatically generated in the correct uniterm sequence.

New documents would be assigned free uniterms, and these would be edited against the uniterm dictionary. On a biweekly basis, the descriptor (edited uniterms) cards accumulated as a result of new cataloging would be consolidated, and correctly sequenced for introduction into the master search term file. These consolidated additions and changes to the search file would be reproduced and sent to each Regional Office for incorporation into their master search files.

The punch card file would be arranged in "Search Term Decks." It is estimated that there would have been 10,000 search term decks (one for each unique uniterm) in the complete file when the system was fully implemented. A very slow growth rate in new terms was predicted.

Each punch card of a search term deck would contain the search term (uniterm - probably in numeric code) followed by a maximum of eleven ASTIA document numbers to which the search terms had been assigned. Where more than eleven AD numbers were included in a single search term record, additional cards would be utilized to make up the "search deck." On the basis of this card capacity, approximately 160,000 individual punch cards would have been required for the 175,000 reports then in the AD collection. The card file would expand at a rate of approximately 1,200 a day. Periodically, the entire file would have to have been consolidated on magnetic tape and repunched from tape-to-cards to integrate new and revised data and to delete obsolete data.

The system would have functioned as follows: The customer would make his needs known, and ASTIA reference personnel would select appropriate search terms. An index search would be initiated to determine those document numbers which satisfied the selected descriptive terms. The search would be accomplished by manually pulling the three or four search term decks and inserting these into the Special Index Analyzer where the desired association of document numbers would be accomplished in a matter of a few minutes. The output would be a list of pertinent document numbers. An on-the-spot validation of the customer's need-to-know would be accomplished, and the customer would be given the abstracts to

review. The customer might also review microcards or microfilms of the documents selected. Full-size copies of the documents would be requested from the central facility.

Under this concept of operation, reserve capability for both descriptor search and validation would have been maintained at the central facility to assist the Regional Offices during peak loads by preparing a list of pertinent validated document numbers in advance of the arrival of a customer when the pertinent data were furnished prior to the customer's visit. (See Figure 1, General Process Flow Chart)

At first the punch card Special Index Analyzer operation appeared to be feasible. However, a more detailed examination of the system cast considerable doubt. The equipment would have to be purchased before any experimentation or operating experience could be gained. The resources required to maintain, update, and manipulate a punch card file of this size and complexity would in the long run be greater than the manual operation it replaced. The integrity of the file would be a major concern in the operation of the system. Thousands of punch cards would be pulled from the files to support the daily operations, and then refiled. These cards, plus cards for new reports being added to the collection, would have to be filed manually on a daily basis. A continuing program for replacing worn out or mutilated cards

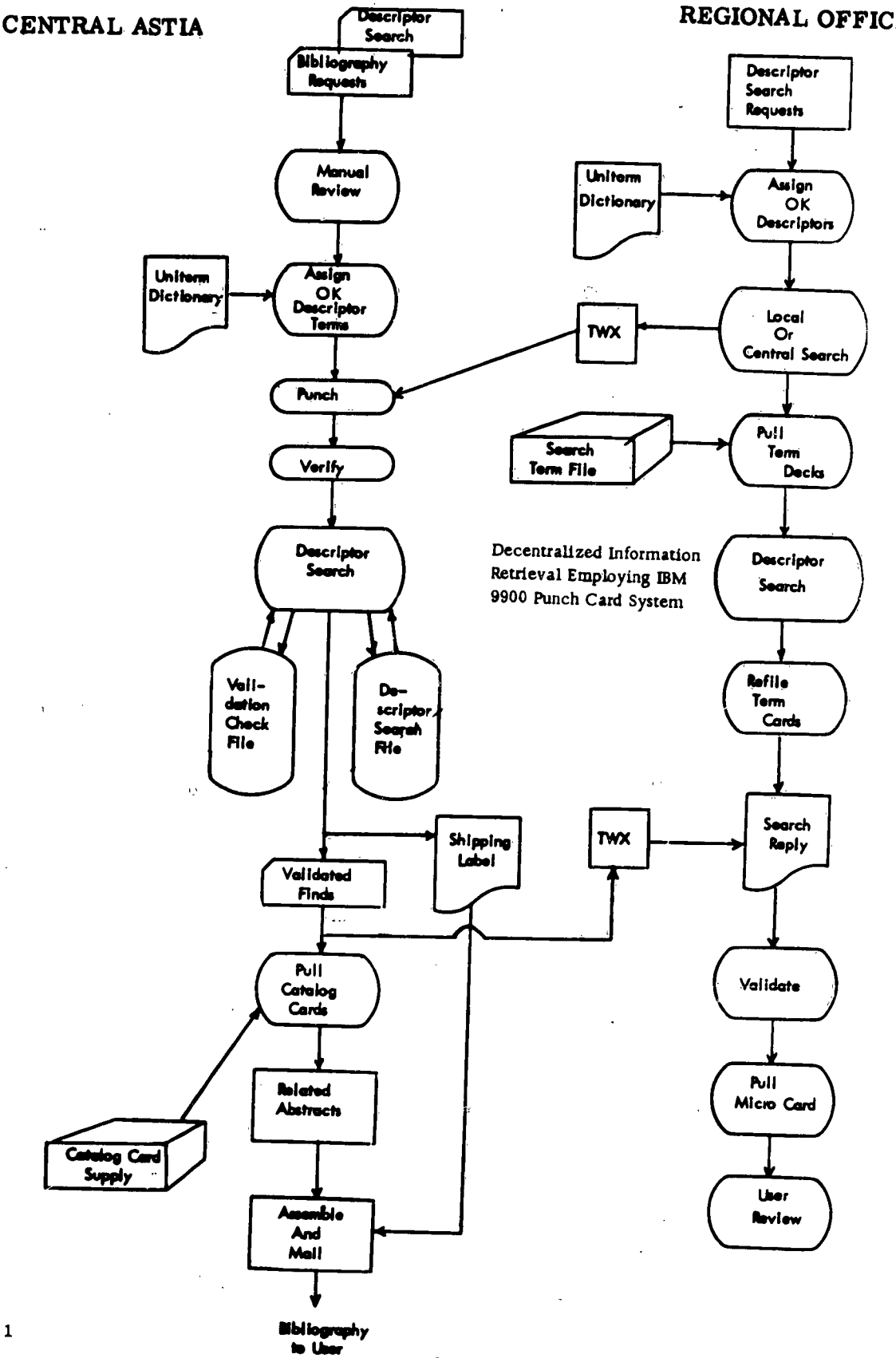
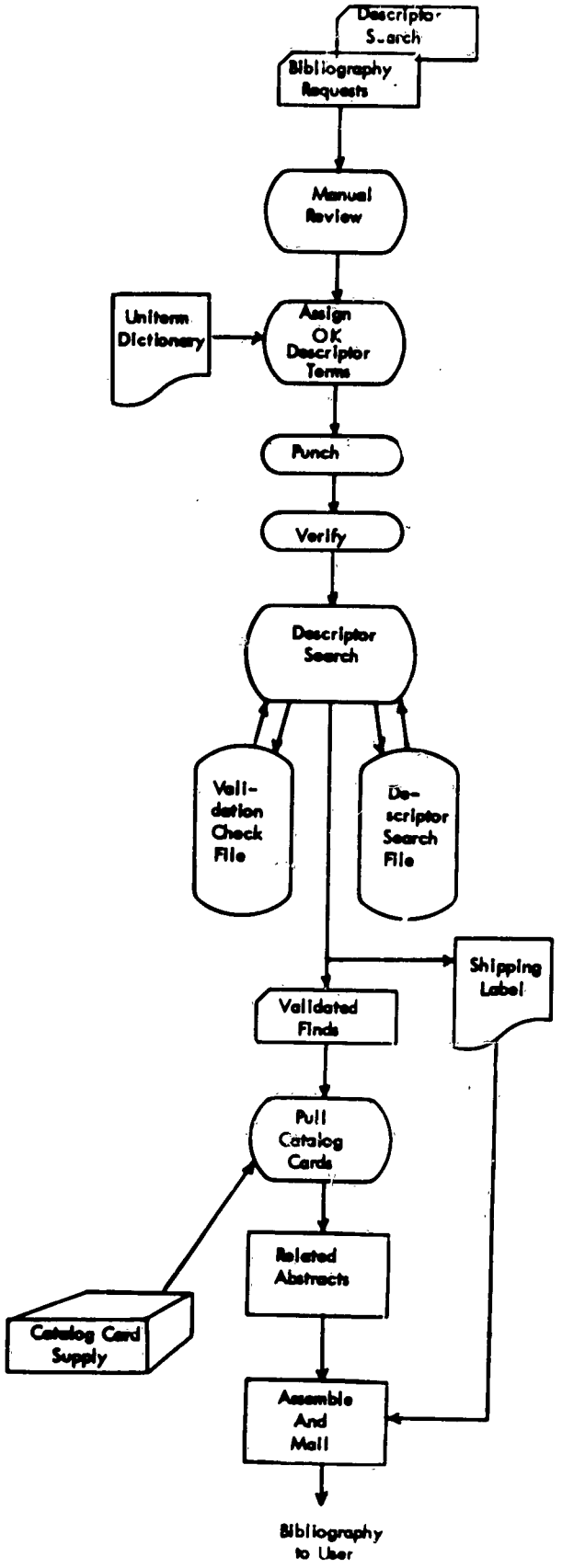

CENTRAL ASTIA
REGIONAL OFFICE
Descriptor Search
Bibliography Requests
Manual Review
Uniterm Dictionary
Assign OK Descriptor Terms
Punch
Verify
Descriptor Search
Validation Check File
Descriptor Search File
Validated Finds
Shipping Label
Pull Catalog Cards
Related Abstracts
Catalog Card Supply
Assemble And Mail
Bibliography to User
Descriptor Search Requests
Uniterm Dictionary
Assign OK Descriptors
Local Or Central Search
TWX
Search Term File
Pull Term Decks
Descriptor Search
Refile Term Cards
TWX
Search Reply
Validate
Pull Micro Card
User Review
Decentralized Information Retrieval Employing IBM 9900 Punch Card System

Fig. 1

would be required. A requirement for periodic reproduction of the entire file in quadruplicate was recognized in the original planning.

By the time ASTIA's automation proposal was formalized in November of 1958, the proposal to employ the punch card Special Index Analyzer retrieval system at the Regional Offices had been rejected for the same reason the punch card information retrieval proposals had been rejected in the past. The ASTIA collection of documents was just too large and too diversified for a punch card operation.

The proposal submitted to higher headquarters on 20 November 1958, was to automate request processing, cumulative indexing of ASTIA's Technical Abstract Bulletin (TAB) and document inventory management. The Commander of ASTIA also proposed that one Special Index Analyzer be purchased for the central facility to be operated on an extended experimental basis. A limited number of machine compiled bibliographies would be produced and evaluated during the experimental phase. (See Figure 2, General Process Chart)



Centralized Information Retrieval Employing IBM 9900 Punch Card System

Fig. 2

INITIAL PLANS FOR EMPLOYMENT OF A GENERAL PURPOSE ELECTRONIC COMPUTER FOR INFORMATION RETRIEVAL

In addition to IBM, other manufacturers of data processing equipment made proposals for automation of ASTIA. Remington Rand submitted a proposal for the UNIVAC I which was immediately available and for the new UNIVAC Solid State 90 (USS 90) which could be made available in the card configuration in sufficient time to become operational by 15 February 1960. The magnetic tape configuration would become available in July 1960 and a random access capability was planned for late 1960.

Further study of the Remington Rand UNIVAC proposal for automation of the ASTIA operation led to the conclusion that it would be feasible to employ a general-purpose magnetic tape data processing system for information search and retrieval. By assigning numeric codes to the retrieval terms to be employed in the system, the manipulation of the terms would be greatly simplified. The general systems design and operational concepts for the present ASTIA system of automated information search and retrieval were developed along these lines.

The proposal for automation was modified to reflect the newly formulated concepts and detailed planning began. The USS 90 data processing system was selected and implementation followed closely on the heels of planning. Objectives were defined, tasks were assigned and target dates were established. At that time, the over-all objective

for information retrieval was to establish and operate a system specifically designed for automatic identification of reports in the ASTIA collection by ASTIA document (AD) number on any specified subject or combination of subjects.

With the document numbers identified, catalog cards would be pulled from the files manually and furnished to the customer. There were by this time 200,000 reports in the AD collection representing the documentation of the previous seven years' military research and development accomplishments.

In the initial planning, it was intended to contract with a commercial organization to develop a thesaurus of authoritative descriptors utilizing the assigned uniterm and then reassign the standard terms to the documents in the AD collection. Those retrieval terms not included in the thesaurus as authoritative descriptors were to be retained and employed as "open-ended" or non-controlled retrieval terms. Such terms include project names, geographic locations, etc. Two companies showed interest in the project, but never submitted proposals or indicated a desire to bid.

Consultants in the field of information retrieval advised that the professional personnel best qualified to compile the thesaurus were the same ASTIA personnel who had processed the collection of documents into the system. It soon became apparent that the assignment of retrieval terms to the documents also had to be accomplished in-house.

By 1 September 1959, the compilation of the thesaurus had reached the stage where the assignment of retrieval terms could begin for documents in the existing AD collection and for new material being processed into the system. These two tasks are described in AD-227 000, Automation of ASTIA - A Preliminary Report, 1 December 1959, and AD-247 000, Automation of ASTIA - 1960.

One large conversion task still remained -- the creation of the machine retrieval files. This was to be accomplished by a contractor. On 1 October 1959, the work statement for the key punching of the retrieval terms was passed to the contracting officer, with a 1 December deadline for the work to begin.

In the original approach, it was intended to key punch a card for each term assigned to a document and then sort into alphabetic sequence for data clean-up and subsequent coding in numeric sequence by the computer.

It soon became apparent that if the 1 July 1960 date for implementation of the system was to be met, a new approach was needed. After further study, a decision was reached to prepare the master coding for the authoritative list of descriptors before they had been compiled into the thesaurus. A seven-digit numeric code was assigned to each descriptor so that numeric sequence and alphabetic sequence were coincident. This code was adopted to simplify the machine manipulation of the file. To allow for elasticity, a gap

of 1,000 numbers was left between consecutive descriptors to provide for anticipated expansion and still retain the alpha-numeric relationship.

A closer look at the implementation schedule indicated an increase in tempo was still a necessity. The document processing personnel of ASTIA assumed the added task of assigning the numeric codes to the retrieval terms concurrent with the assignment of the terms to documents. In addition to the numeric descriptor code, a key was added to indicate the terms to be used for subject accumulation to create a machine capability to compile a subject index of the total AD collection.

The work statement was recalled from the contracting officer and modified to reflect the requirement for key punching the numeric codes rather than the actual descriptors. Although this delayed the awarding of the contract 15 days -- from 1 December to 15 December 1959 -- it actually placed ASTIA further ahead in implementation and also resulted in a considerable reduction in cost by greatly reducing the alpha key punching required to convert the file. On a competitive bidding basis, the contract was awarded to the Service Bureau Corporation, a subsidiary of IBM.

The in-house assignment of retrieval terms began slowly because of the additional requirement to assign the numeric codes concurrently. In spite of the slow start, it was soon recognized that the decision

had been a wise one. The magnetic tape system did not become available in early March as anticipated.* Without the added capacity of the tape system, three months of punch card processing would have been required to arrange the data for operational runs. This punch card processing could not have begun until the conversion contract was completed on 24 June 1960. Thus, several months' head start was gained by punching the numeric codes for the retrieval terms into the original machine records. It also made possible extensive experimentation and actual production of bibliographies on the card computer several months in advance of the magnetic tape operation.

The assignment of retrieval terms and the key punching of the file increased in tempo. Both operations were soon back on the original schedule. The last of the data to be key punched was completed by the contractor on 24 June 1960, and thereafter ASTIA assumed full responsibility for maintaining and updating the file. By this time, more than a million and a quarter punch card records had been prepared. A sample of each of the three types of punch card formats is shown in Figure 3.

* The installation of the USS 90 tape system was to be on an operational test basis until the scheduled delivery date of 1 July 1960.

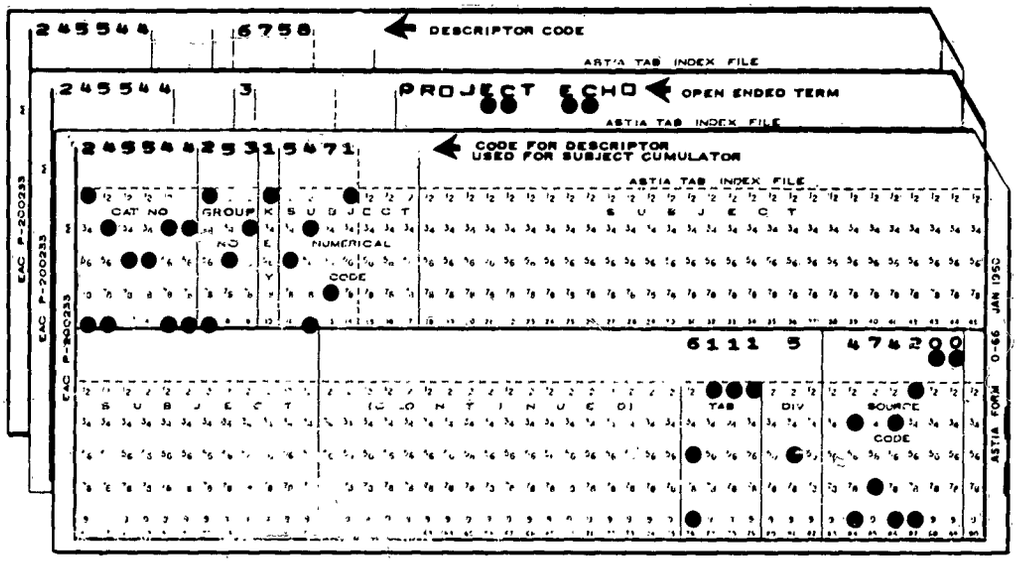

245544 6758
DESCRIPTOR CODE
ASTIA TAB INDEX FILE
245544 3
PROJECT ECHO
OPEN ENDED TERM
ASTIA TAB INDEX FILE
245544 25315471
CODE FOR DESCRIPTOR USED FOR SUBJECT CUMULATOR
ASTIA TAB INDEX FILE
CAT NO
GROUP NO
KEY
SUBJECT
NUMERICAL CODE
SUBJECT
6111 5
474200
SUBJECT
CONTINUED
TAB
DIV
SOURCE CODE
EAC P-200233
EAC P-200233
EAC P-200233
ASTIA FORM 0-66 JAN 1950

Fig. 3

EXPERIMENTATION IN AUTOMATED INFORMATION RETRIEVAL

Experimentation in machine retrieval progressed rather slowly at first because insufficient data were available for valid testing. The original system was a take-off from a manual operation that had withstood the test of time, and this lack of extensive theoretical testing was not considered too critical. The priority for assignment of terms and for key punching was shifted to the AD collection in the 200,000 and 300,000 series (material cataloged subsequent to August 1958) so that testing and limited service on newly acquired documents could be combined. By early May of 1960, the punch cards for this group of documents were reproduced and sorted into retrieval term decks for machine manipulation. The file was updated as new documents were added to the collection.

Originally it had been planned to experiment with machine retrieval on the punch card equipment. The Remington Rand consultants had recommended their latest punch card numerical collator 319.1 be employed for retrieval experimentation and limited production until their USS 90 magnetic tape system became available. However, due to the mass of card files required for extensive testing--some 300,000 cards--it was decided to utilize the UNIVAC Solid State card system for the testing and to write programs that would allow the descriptor decks to be introduced in random sequence,

thus eliminating the need for extensive off-line sorting. The experimentation was to be limited to 26 test bibliographies that had been previously processed manually. A diagram of the test program is shown in Figure 4.

The program originally developed for the experimentation on the punch card system was based on an up-to-date examination of customers' requests for bibliographic service. Close attention was also given to the character of the retrieval term file being created for the mechanized operation.

A working-level study group, headed by Joseph M. Powers of the ASTIA Washington Reference Center, was established to conduct experimentation and develop operational procedures for machine compilation of bibliographies. The specific objective of this group was to formulate guidelines for selecting retrieval terms and for prescribing the depth and extent of coordination of terms to satisfy bibliography requests.

One of the early experiments conducted by the group was to have the scientific analysis personnel, who assign the retrieval terms to documents, and the reference personnel, who compile bibliographies manually, separately select the retrieval term patterns for computer searches to satisfy 20 bibliography requests that had been received by ASTIA. Not too much was expected from a single experiment of this nature; however, the results as in this

GENERAL PROCESS CHART

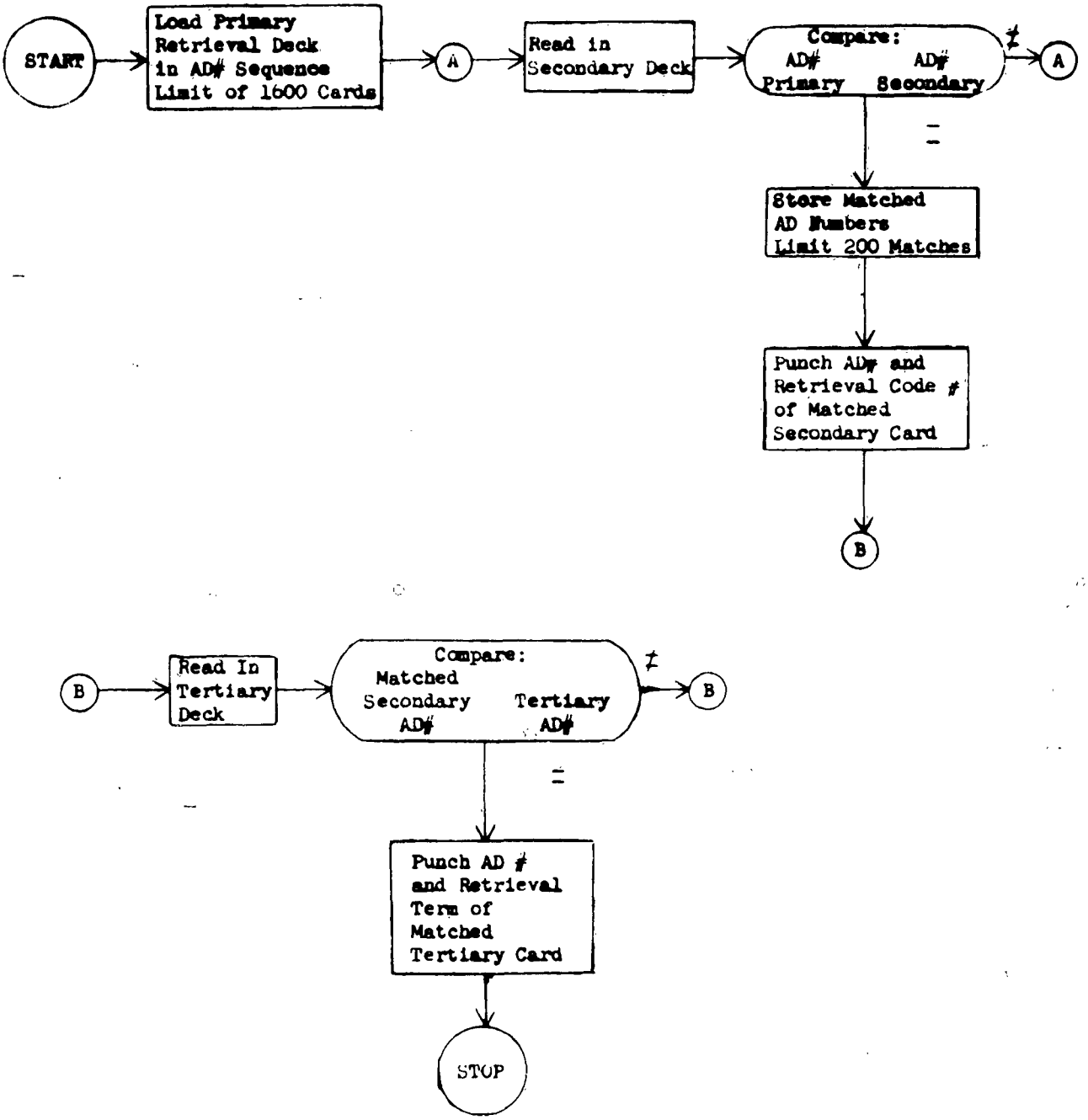


Fig. 4 Information Retrieval in Card Operative System

instance were often enlightening. The bibliography and reference group selected a total of 366 retrieval terms to satisfy the 20 searches. For a single bibliography, this group ranged from a low of four retrieval terms to a high of 47, and an average of 18 per bibliography.

The Scientific Analysis personnel selected a total of 363 terms, three less than the first group's total, for an average of 17 terms per bibliography. The spread of retrieval terms selected by the second group to compile the 20 bibliographies was even greater than the first group. It ran from a low of three to a high of 61. In four instances out of twenty, each group selected an identical number of terms.

The exercise just described was typical of many carried out over a period of several months. Each experiment in itself provided little more than another new bit of information. These new bits, however, began to fit into a pattern. Only in one instance was extensive statistical sampling of the retrieval files possible. Mrs. Claire Schultz, of the Applied Mathematics Department of Remington Rand UNIVAC who worked with ASTIA in the information retrieval area, provided some very valuable statistics on the frequency distribution of retrieval terms through machine manipulation of the retrieval term files for 38,000 documents.

PUNCH CARD RETRIEVAL SYSTEM ON THE UNIVAC SOLID STATE 90 COMPUTER

The programming for the experimentation in machine retrieval discussed earlier was prepared for a one-time effort to compile 26 bibliographies to test the retrieval system.

During May and June of 1960, the requests for report bibliographies far exceeded the manual resources. The Remington Rand Corporation, with ASTIA's consent, had readjusted the magnetic tape delivery schedule to allow for factory engineering of the tape drives to a new console. A second USS 90 card system was made available from 15 June through 7 December so that work would not be interrupted during installation and conversion to the new system. These events, plus the favorable results of the initial tests, led to a decision in mid-June to go into full production of mechanized information retrieval under the card operation. The production would be limited to the last two years of documentation.

The program for experimentation in machine retrieval was designed to provide a maximum of flexibility; little attention was given to program efficiency. When the decision was made to utilize the experimental program for actual production, modifications were made to improve the efficiency. The capacity for employment of the primary descriptor deck was increased and a magnitude compared to the last

item of each memory band was added to limit the search to those bands that contained data within the range of the secondary and tertiary decks. The program was designed to coordinate groups of retrieval terms as well as individual terms. The group of primary term decks were sorted together in AD sequence and stored in memory. A total of 1,600 records could be accommodated in the primary deck for a single run. If the number were greater, an additional computer run was necessary. The secondary group of term decks were fed into the card reader in random sequence. There was no program limit to the number of records employed in the secondary decks.

When the computer identified a match between an AD number in the primary group stored in memory and secondary group in the card reader, the common AD number was stored in memory, and a card was punched out to identify the matching AD number and the code of the secondary descriptor. When the third level, or tertiary group of terms, was introduced through the card reader, comparison was made only between AD numbers of the tertiary term decks and those AD numbers stored in memory as a result of primary and secondary match. Again the output was in the form of punch cards, showing the AD number common to all three groups and the tertiary descriptor code.

During the month of July 1960, 269 bibliographies were compiled on the USS 90 card system.

Although this computer approach has a great deal of flexibility -- a simulation of random access through manual manipulation of the punch card search decks -- it has the same major deficiency that caused ASTIA and its predecessors to abandon the punch card approach in the past. Because of the large volume of punch cards involved, the file updating and maintenance tasks were so great that only 20% of the ASTIA collection could be employed.

PROGRAMMING CONCEPTS - MAGNETIC TAPE

The original target date for machine retrieval on the magnetic tape system had been established for 1 July 1960. Under this deadline, it became necessary to begin programming for the tape operation long before any experimental or test data could be accumulated.

The initial programming efforts had to be based on a strictly theoretical concept for retrieval through a coordination of individual retrieval terms to the depth required to isolate reports on a given subject. The first USS 90 tape program for retrieval was written under these circumstances by several Remington Rand UNIVAC and ASTIA programmers, under the immediate supervision of John K. Henderson of the Remington Rand UNIVAC Federal Government Sales Office. A block diagram of the program is shown in Figure 5.

Debugging and testing this program was discontinued when the delivery date for the tape system was rescheduled for August 1960. It had become apparent that before the tape system could become operational, considerable operational data would become available through experimentation and limited production on the USS 90 punch card system.

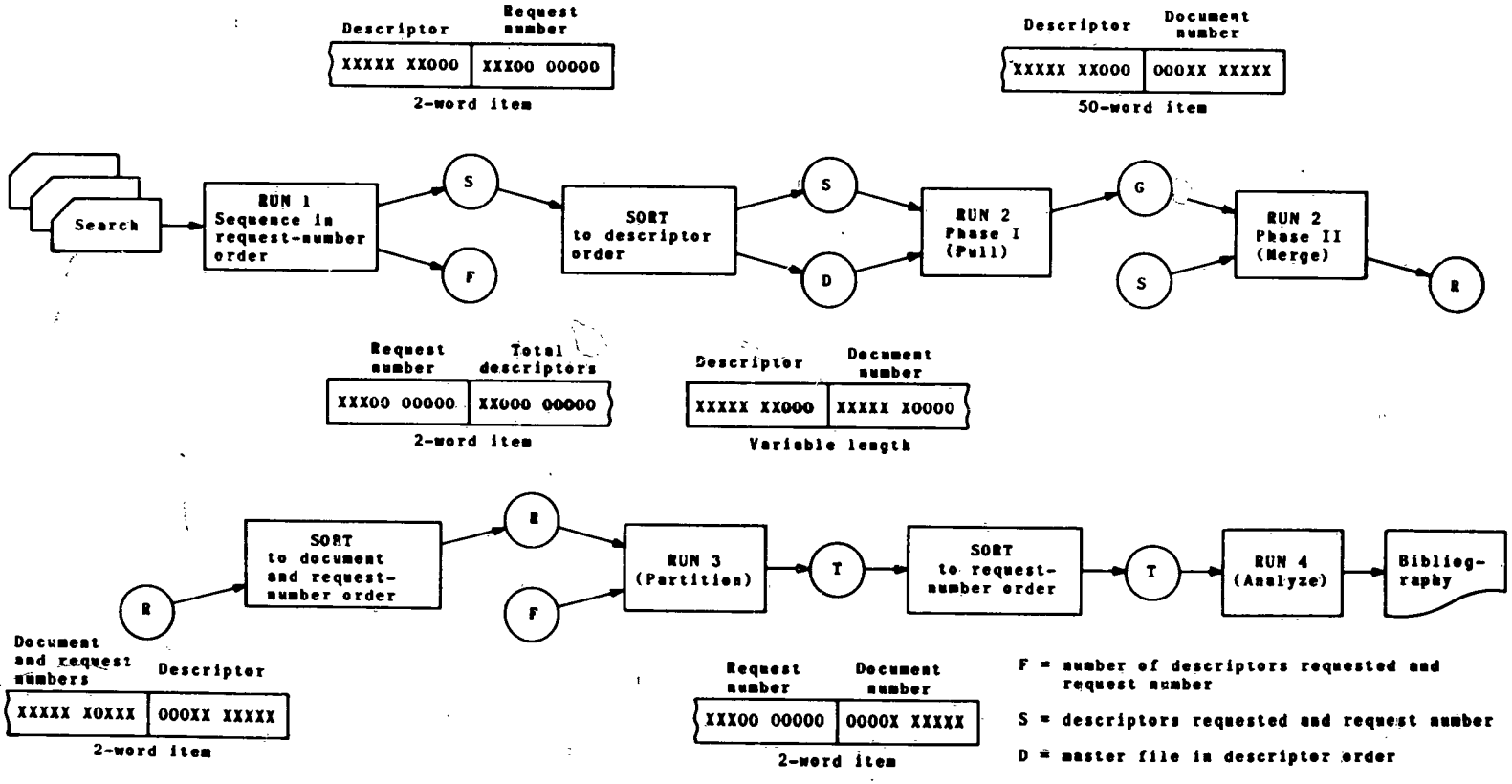


Fig. 5

Block Diagram and Sample Tapes of Information-Retrieval Process

A good deal was learned from the early programming exercises. It was found that the format planned for the master retrieval term tape file -- ASTIA document number sequence, followed by the descriptors assigned to the document -- did not lend itself to efficient programming. A new format for the master file was designed which permitted a new approach to the machine manipulation of the data. This new format contained two 50-word items on each block of tape. The first word of each item is the retrieval term code, and the next 49 words contain AD numbers. If there are less than 49 AD numbers, a "zero" fill is used to complete the 50-word item. When there are more than 49 AD entries to a single retrieval term record, the code is repeated in the first word of each 50-word item required to complete the record. Each block of tape has a 100-word capacity. Thus, there are 98 AD entries possible in a single block or 470,400 to a single 2,400 foot reel of magnetic tape. The current master retrieval term record for the AD collection requires four tapes. A printout of the tape format of the master file is shown in Figure 6.

Evidence began to build up, particularly from the operational experience with the punch card system, that the early programming concepts for information search and retrieval on the tape system would not meet the requirements as these requirements were being expressed by the customers. Additionally, the increasing customer demands for report bibliographies indicated that before the system

Fig. 6

Data on a Single Block of Tape

Two 50-word items are contained on each block of tape. The first word of each item is the Descriptor code, and the next 49 words are AD numbers. If there are less than 49 AD numbers, a "zero" fill is used to complete the 50-word item.

When there are more than 49 AD's to a single Descriptor record, the Descriptor code is repeated in the first word of each 50-word item required to complete the Descriptor record. Each block of tape has a 100 word capacity. Thus, there are 98 AD entries possible in a single block or 470,400 to a single 2,400 foot reel of magnetic tape. The currect AD collection requires 4 tapes.

could become operational, the workload would exceed the planned capacity. A new approach to the machine program was indicated.

Armed with more factual data than had ever before been available, it was possible to write a new magnetic tape program that would satisfy the immediate needs of the customers and also retain a potential for flexibility and expansion. The program was written by ASTIA personnel. Mr. Hugh Donaghue of Datatrol Corporation, an ASTIA programming consultant, reviewed the computer logic and assisted in writing and debugging the program, particularly the tape control techniques.

The new program was designed to handle as many as ten searches simultaneously with a maximum of four retrieval term coordinations for any search. The ability to combine a group of terms and process them as a single term was retained from the card program. The ability to exclude any term or any group of terms from a search was also written into the program. Actually, the only restriction to the number of terms that may be employed within these limitations is determined by the optimum operating time for a single run. It requires no more time to record a term record for ten searches than for a single search. Thus, the greater the number of common terms employed in simultaneous searches, the more efficient the computer run.

PREPARATION FOR COMPUTER SEARCH BY THE REFERENCE ANALYSTS

It is the responsibility of the reference analysts to determine the retrieval terms to be employed and the depth of coordination required for the computer search program. He must determine what information is required to satisfy a search and how the information has been cataloged into the system.

How well these two factors are determined and translated into the search program, largely depends on the reference analysts' judgment and experience. However, the outcome is greatly influenced by the way the retrieval terms were assigned to the documents as they were processed into the system. The document analyst in cataloging documents must tag them by a set of retrieval terms both specifically and generically. He must also build relationship between assigned terms which to some extent will anticipate inquiries and provide for retrieval from a variety of subject approaches.

In determining what information is required, the reference analyst must interpret the question, define its limits, and usually access its unwritten applications. The adequacy of the descriptive data provided by the requester, to a great extent will influence the specificity of the search results.

An example of an actual request for information processed by ASTIA is shown on the following pages. The bibliography request form and the internal processing form shown in Figures 7 and 8 are in current use.

It may be noted in the example that groups of descriptors have been combined and treated as a single descriptor for the computer search. In this particular search there are three separate considerations. The first concerns personnel and the radiation effects or damage caused by microwaves. The second concerns the effectiveness of shielding against microwaves; the third concerns more general reports on the measurement and analysis of radiation effects and hazards occurring in underground structures.

BIBLIOGRAPHY REQUEST		Date: February 14, 1961
1. TO: ASTIA (TISBB) Arlington Hall Sta Arlington 12, Va	2. FROM: Jones & Smith, Inc. 4500 Bradbury Blvd. Denver, Colorado	3. ASTIA Code No.: 7654321 4. Requester's Reference: G M 72 5. Requester's Telephone No.: DE 2-2100

6. SUBJECT OF BIBLIOGRAPHY. *Describe in your own words exactly what you want. Explain any special aspects to be covered and any limitations on the scope of the bibliography. Background information, such as TAB references, contract numbers or relevant subjects, may be used to help define your request.* (CAUTION: Classify this form when classified information is included.)

Provide information on the subject of radiation injuries caused by electromagnetic waves in the microwave frequency spectrum, and particularly hazards associated with radar equipment in underground tracking or launch facilities. Also indicate the shielding required for personnel protection.

7. Time coverage desired
☐ Past year ☐ 3 years ☒ 5 years ☐ other

8. Highest security classification of references to be included
☐ Unclassified ☐ Confidential ☒ Secret

9. Request submitted for

a. Contractors must complete this section	b. Government agencies must complete this section
Contract No. AF 33(616)1234	Official Government Use
Signature & Title: James Connelly, Project Engineer	Signature & Title

10. FOR ASTIA USE ONLY

ARB No.	Date Received
ARB 5812	16 Feb 61

ASTIA Form Mar 60 4 Replaces ASTIA Form 4, Sep 58, and ASTIA Form 14, Jan 59, which will be used until stocks are exhausted.

Fig. 7

ASTIA BIBLIOGRAPHY SEARCH RECORD

ARB 5812 Search 1 of 1 Classified Secret AD 70,000 Librarian RBH

Requestor Jones & Smith, Inc., 4500 Bradbury Blvd., Denver, Colo.

Subject Microwave Hazards

Retrieval Terms / Schedules / Open-end Terms

	(1)		(2)		(3)	
	Electromag. Waves	1974	Hazards	2788	Personnel	4500
	Microwaves	3824	Radiation Damage	5059	G. M. Personnel	2710
( )	Microwave Frequency	3818	Radiation Hazards	5061	Radar Operators	5032
	Ultrahigh Frequency	6557.5	Radiation Injuries	5062	Tissues (Biology)	6335
	Radar	5008	Elec. Shielding	1971	Effectiveness	1886
	Radar Equipment	5023	Shielding	5588		
( )						
( )						
)						

Fig. 8

DESCRIPTION OF THE COMPUTER SEARCH

The input to the system is on punch cards. One card as shown in Figure 9 is key punched for each retrieval term prescribed by the reference analyst for each search. These cards are then combined with the input cards for other searches scheduled for simultaneous computer processing.

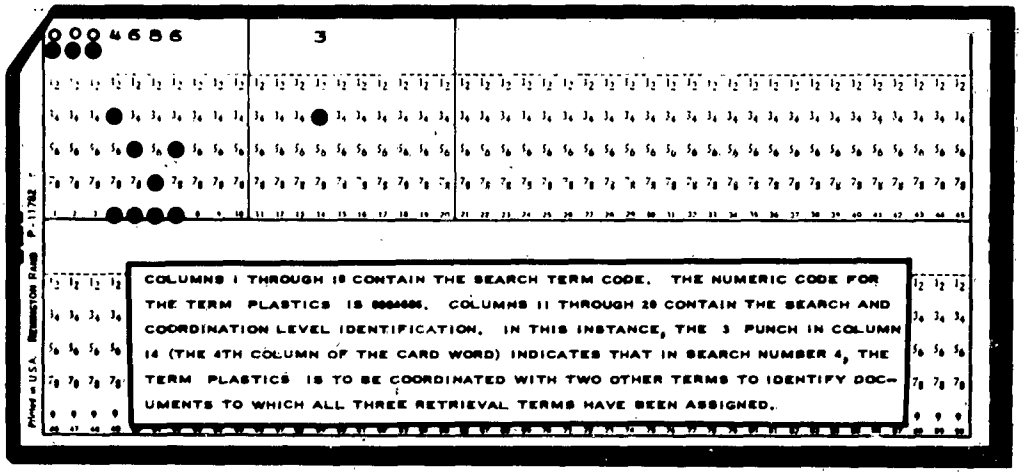


Fig. 9 SEARCH AND RETRIEVAL INPUT CARD

The input cards are sorted off-line to retrieval term code sequence. As the cards are read into the system, cards with like codes are combined into a single two-word record. The first word contains the retrieval term code and the second contains a consolidation of the search and coordination level data. The master

retrieval term tape file is read until a match with the term code on the input card is obtained.

Under program control, the master record for this retrieval term is copied on Output Tape I in a two-word record, with 50 such records to a tape block. The first word of each record is the AD number from the master file. The second word of each record is the search and coordination level identification data as consolidated from the input cards.

The next card input is read, and the matching master record is merged on Output Tape II in the two-word record with the data previously copied on Output Tape I. In the merge, when matching AD numbers are found, the AD number is written on the new output in the first word, and the search and coordination level data are consolidated in the second word. This process continues until the end of the search. At that time, the final output tape contains the identification of the document numbers to satisfy the searches.

Under program control, punch cards are prepared from the output tape, indicating the documents that satisfy the searches, the bibliographic identification code, and date of run. A sample output card is shown in Figure 10.

The output cards are segregated on an off-line. As a general policy, a card is punched for all second or greater levels or coordination. The reference analyst has the option of retaining or eliminating

various levels of coordination. This option can be exercised before or after the computer run. The flow charts for the computer program are shown in Figures 11 through 18.

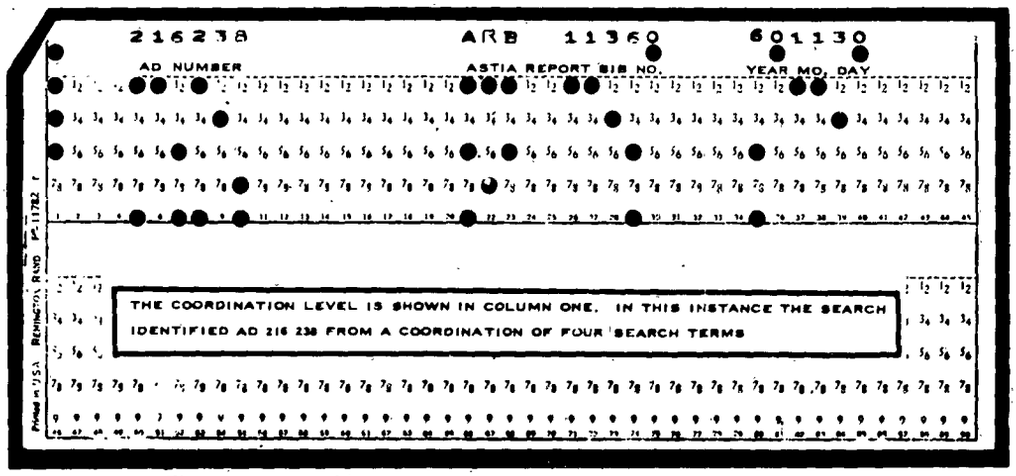
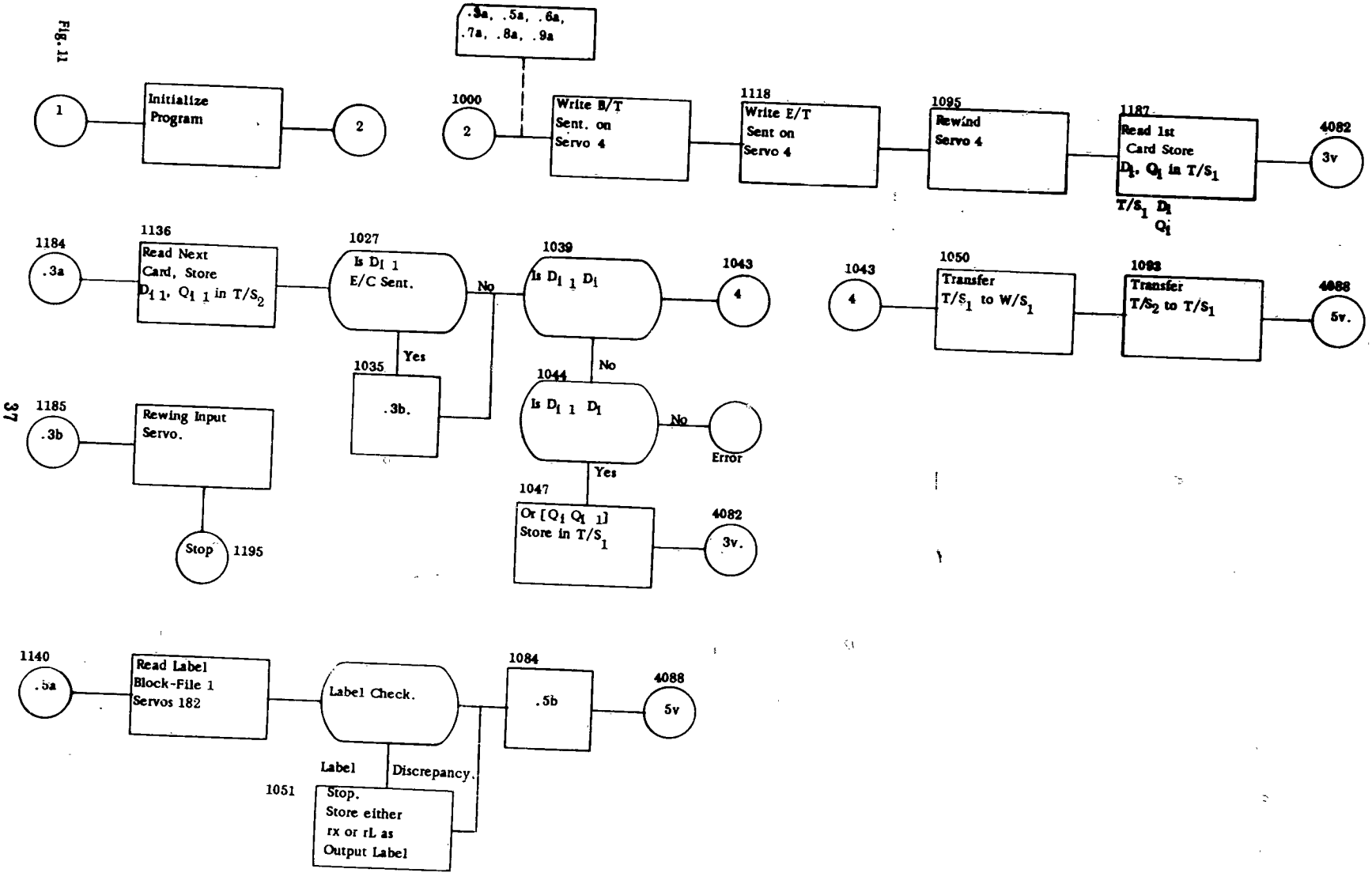


Fig. 10 SEARCH AND RETRIEVAL OUTPUT CARD

Fig. 11


1
Initialize Program
2
.$a, .5a, .6a, .7a, .8a, .9a
1000
2
Write B/T Sent. on Servo 4
1118
Write E/T Sent on Servo 4
1095
Rewind Servo 4
1187
Read 1st Card Store D_1, Q_1 in T/S_1
T/S_1 D_1 Q_1
4082
3v
1184
.3a
1136
Read Next Card, Store $D_{1\,1}$, $Q_{1\,1}$ in T/S_2
1027
Is $D_{1\,1}$ E/C Sent.
No
Yes
1035
.3b.
1039
Is $D_{1\,1}$ D_1
1043
4
1044
Is $D_{1\,1}$ D_1
No
Error
Yes
1047
Or [Q_1 $Q_{1\,1}$] Store in T/S_1
4082
3v.
1185
.3b
Rewing Input Servo.
Stop
1195
1043
4
1050
Transfer T/S_1 to W/S_1
1093
Transfer T/S_2 to T/S_1
4088
5v.
1140
.5a
Read Label Block-File 1 Servos 182
Label Check.
Label
Discrepancy.
1051
Stop. Store either rx or rL as Output Label
1084
.5b
4088
5v

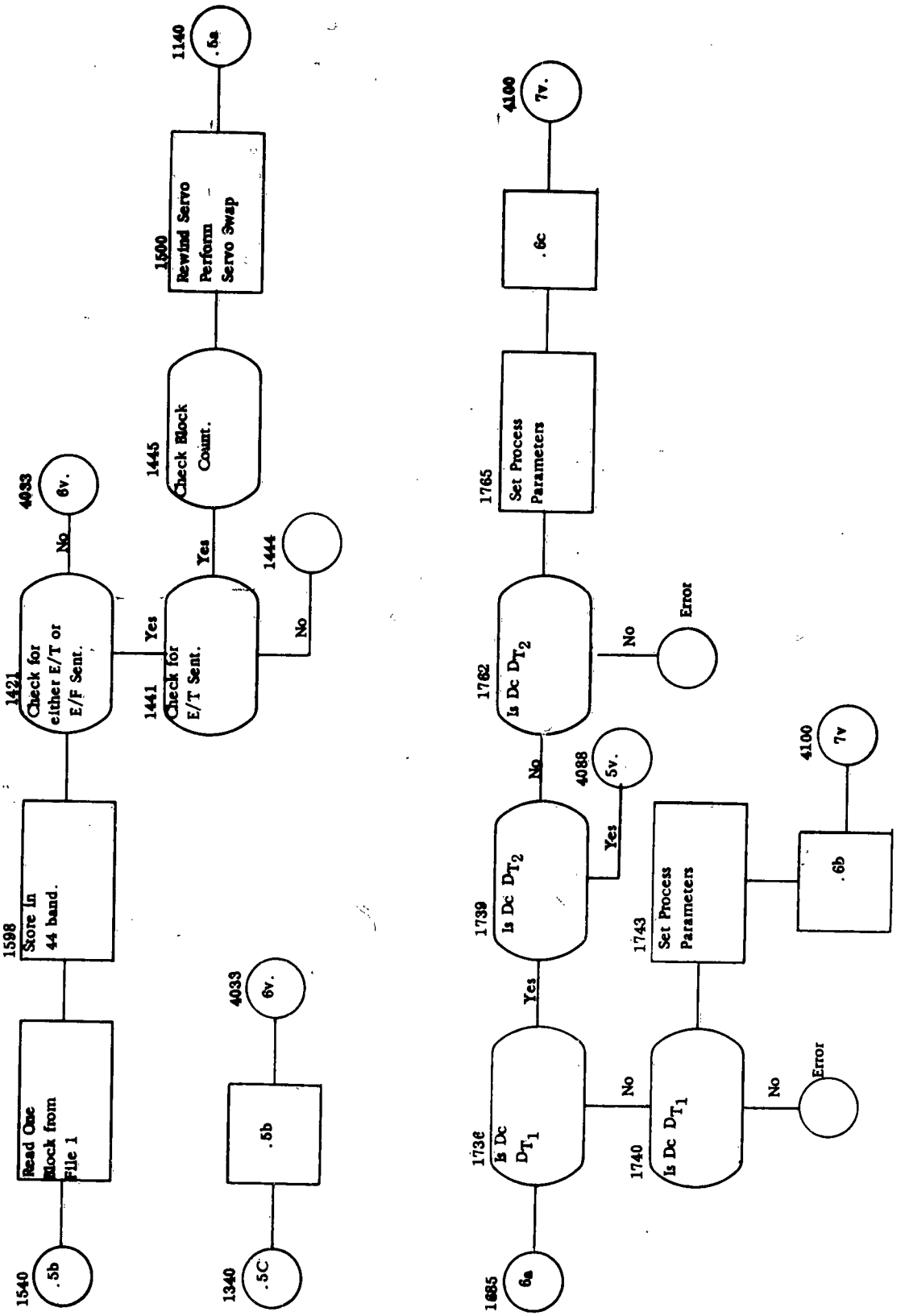

1540
.5b
Read One Block from File 1
1598
Store In 44 band.
1421
Check for either E/T or E/F Sent.
4033
6v.
No
Yes
1441
Check for E/T Sent.
1445
Check Block Count.
Yes
1500
Rewind Servo Perform Servo Swap
1140
.5a
1444
No
4033
6v.
.5b
1340
.5C
1685
6a
1738
Is Dc D_{T_1}
Yes
No
1739
Is Dc D_{T_2}
Yes
No
4088
5v.
1762
Is Dc D_{T_2}
No
Error
1765
Set Process Parameters
4100
7v.
.6c
1740
Is Dc D_{T_1}
No
Error
1743
Set Process Parameters
.6b
4100
7v

Fig. 12

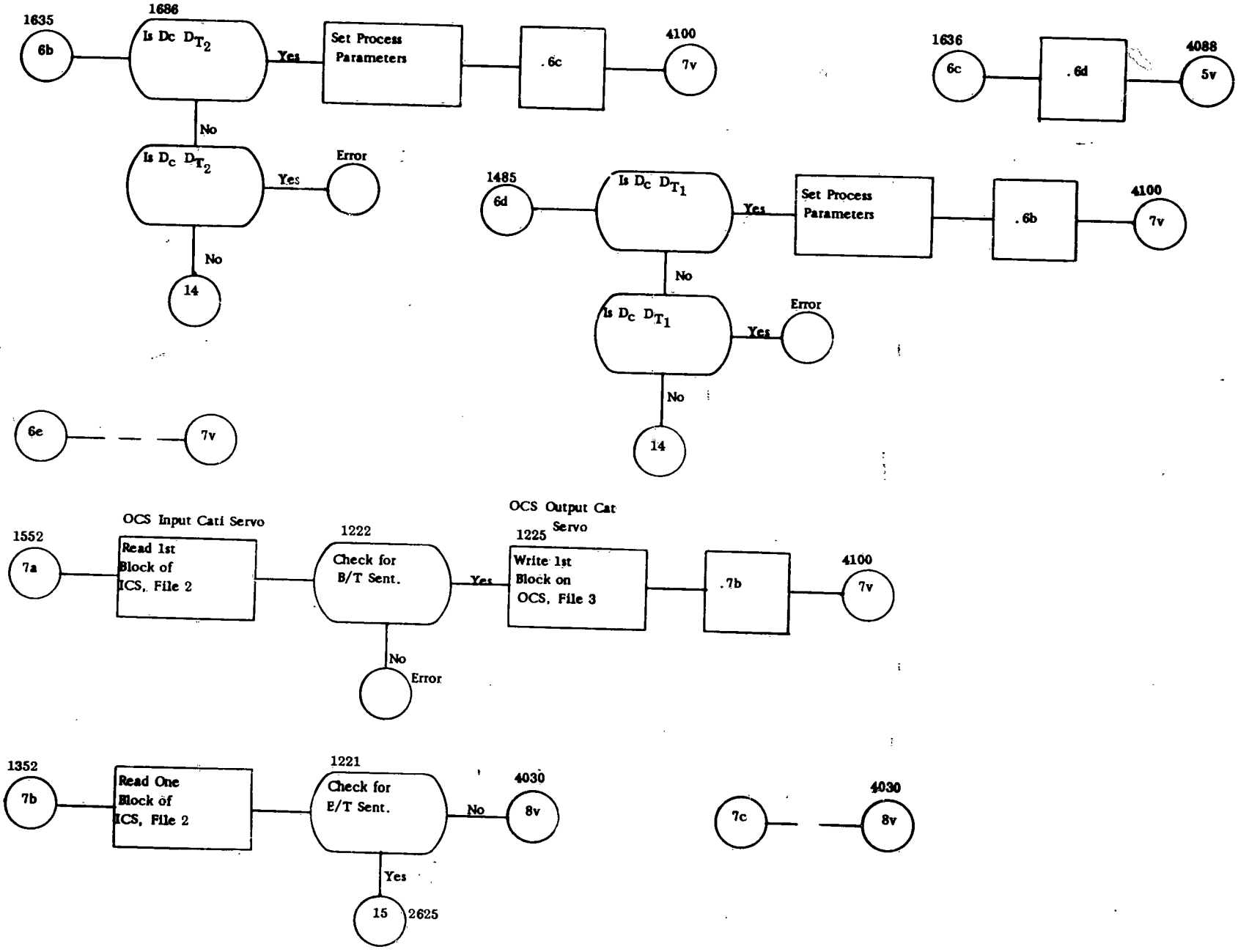
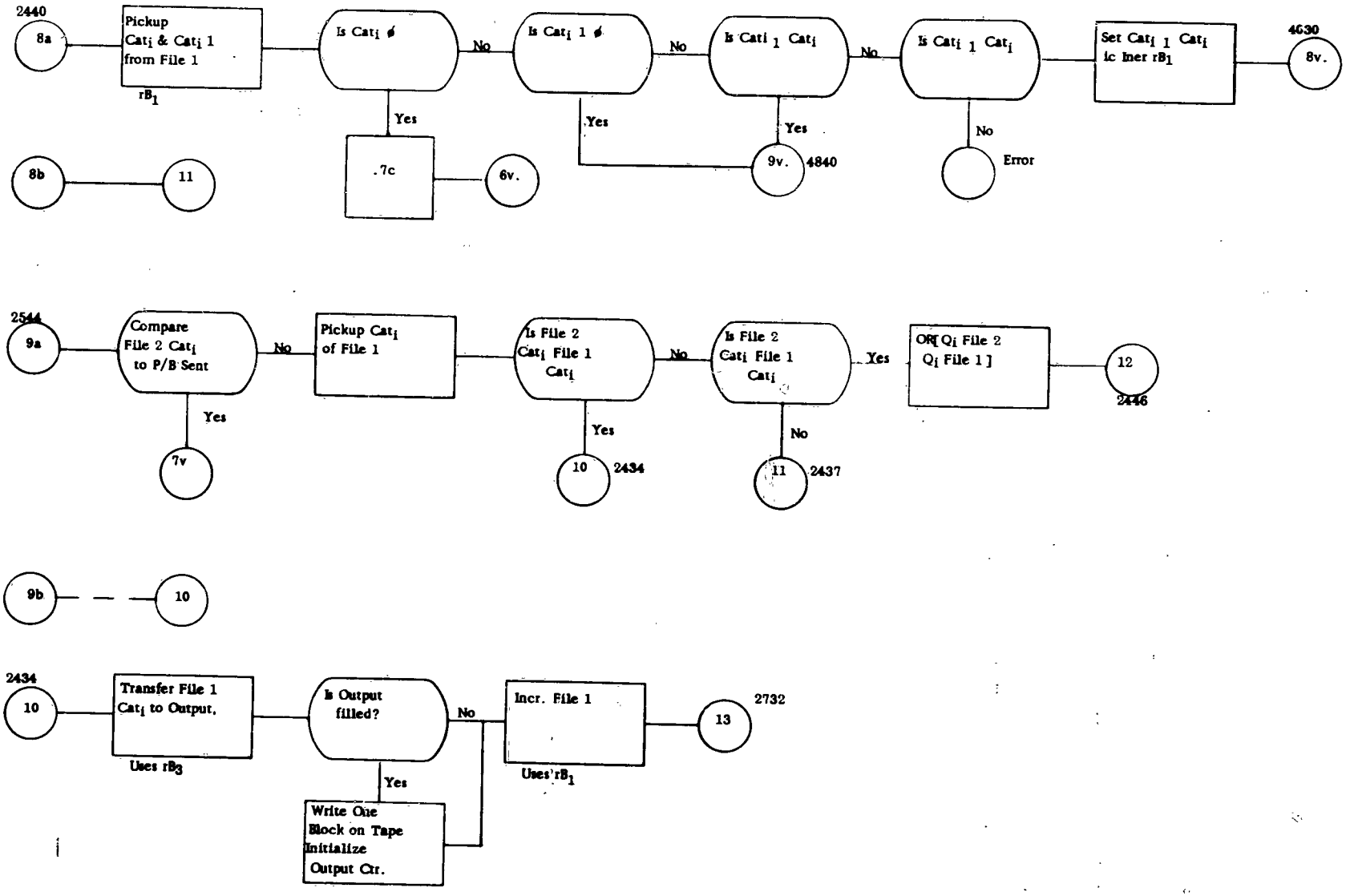

1635
6b
1686
Is Dc D_{T_2}
Yes
Set Process Parameters
.6c
4100
7v
No
Is D_c D_{T_2}
Yes
Error
No
14
1636
6c
.6d
4088
5v
1485
6d
Is D_c D_{T_1}
Yes
Set Process Parameters
.6b
4100
7v
No
Is D_c D_{T_1}
Yes
Error
No
14
6e
7v
OCS Input Cati Servo
1552
7a
Read 1st Block of ICS, File 2
1222
Check for B/T Sent.
Yes
OCS Output Cat Servo
1225
Write 1st Block on OCS, File 3
.7b
4100
7v
No
Error
1352
7b
Read One Block of ICS, File 2
1221
Check for E/T Sent.
No
4030
8v
Yes
15
2625
7c
4030
8v

Fig. 13

Fig. 14


2440
8a
Pickup
Cat$_i$ & Cat$_i$ 1
from File 1
rB$_1$
Is Cat$_i$ ∅
No
Is Cat$_i$ 1 ∅
No
Is Cati$_1$ Cat$_i$
No
Is Cat$_i$ $_1$ Cat$_i$
Set Cat$_i$ $_1$ Cat$_i$
ic Iner rB$_1$
4630
8v.
Yes
Yes
Yes
No
8b
11
.7c
6v.
9v.
4840
Error
2544
9a
Compare
File 2 Cat$_i$
to P/B Sent
No
Pickup Cat$_i$
of File 1
Is File 2
Cat$_i$ File 1
Cat$_i$
No
Is File 2
Cat$_i$ File 1
Cat$_i$
Yes
OR[Q$_i$ File 2
Q$_i$ File 1]
12
2446
Yes
Yes
No
7v
10
2434
11
2437
9b
10
2434
10
Transfer File 1
Cat$_i$ to Output,
Uses rB$_3$
Is Output
filled?
No
Incr. File 1
Uses rB$_1$
2732
13
Yes
Write One
Block on Tape
Initialize
Output Ctr.

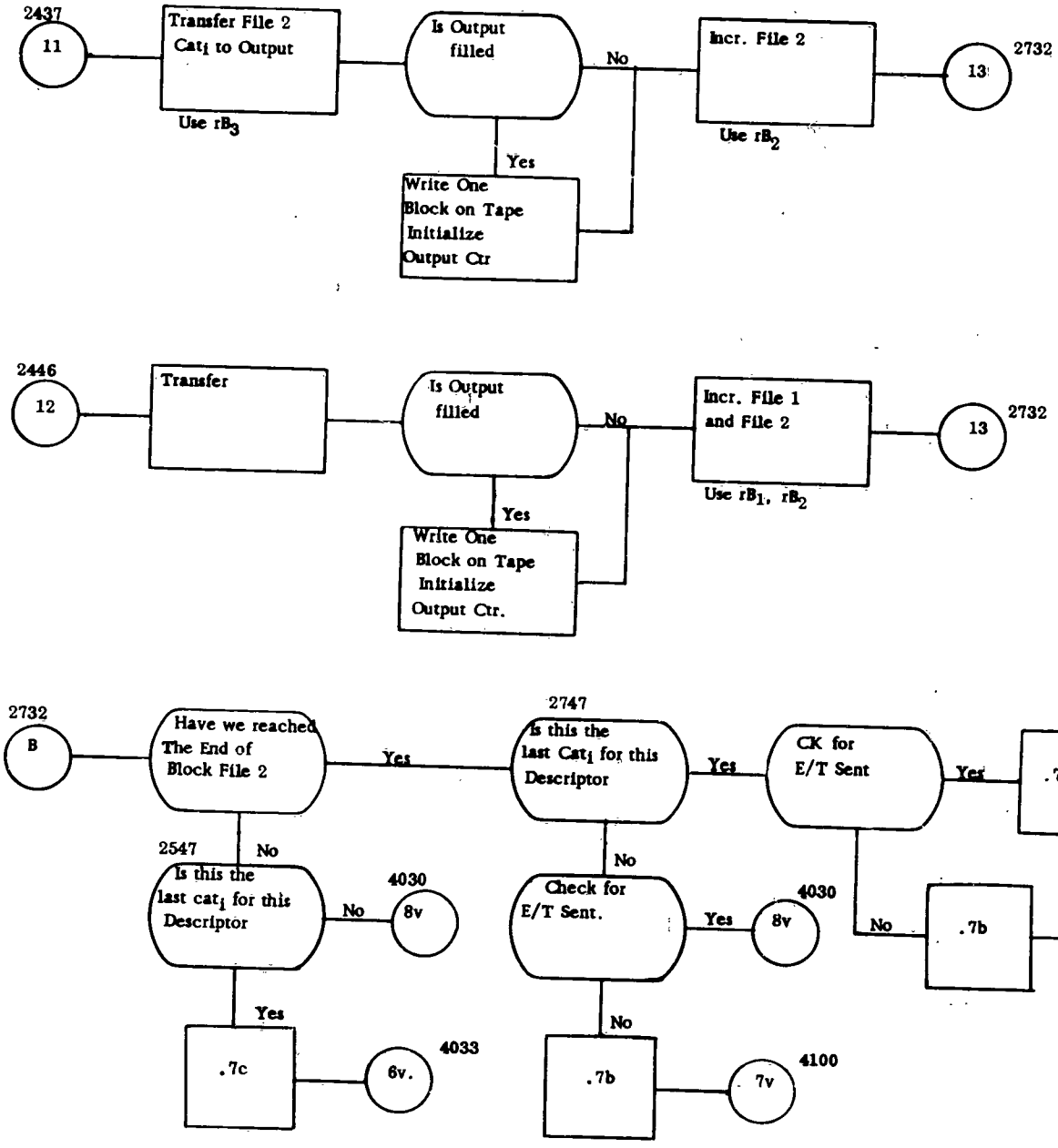
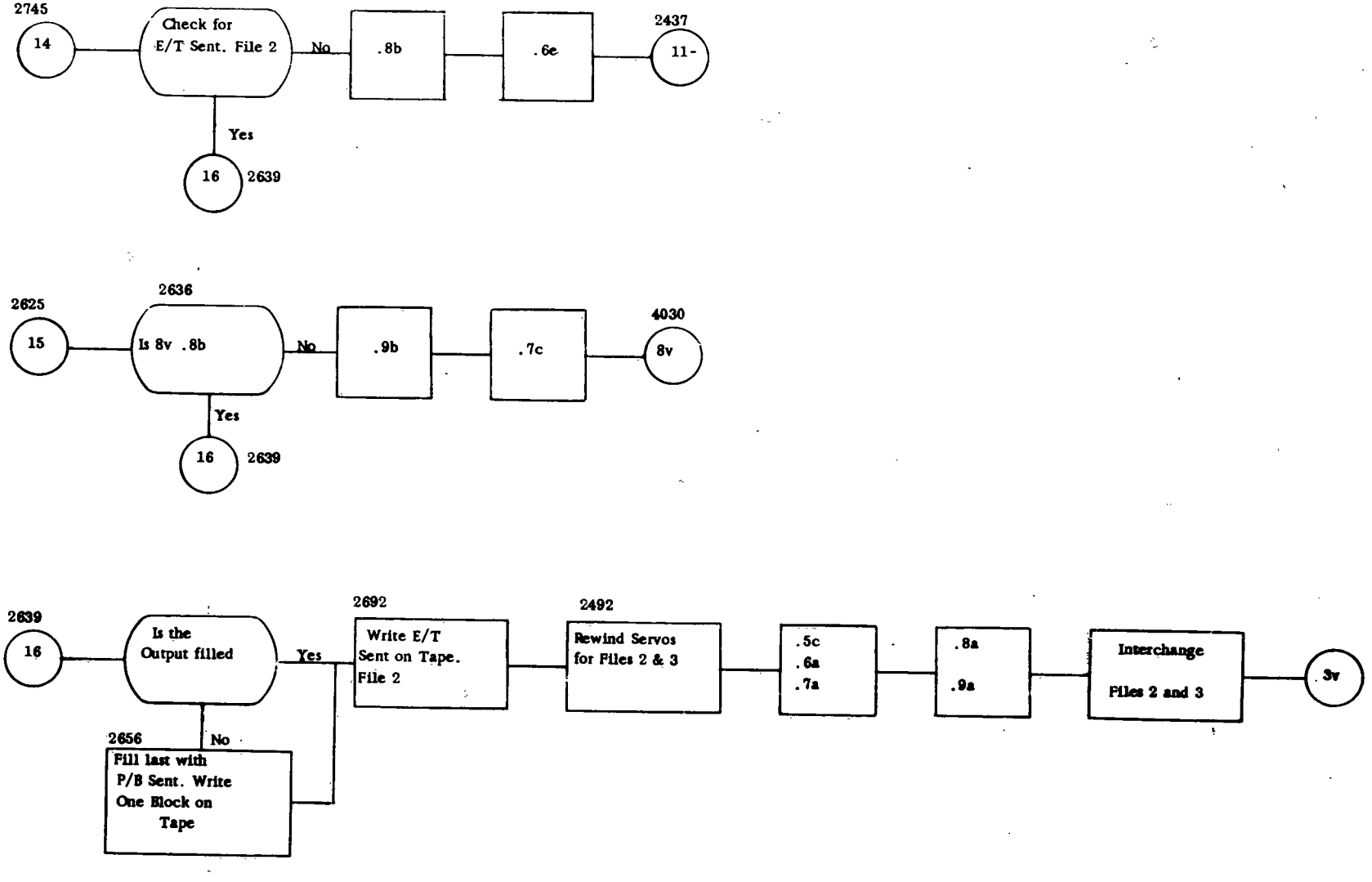

2437
11
Transfer File 2
Cat$_1$ to Output
Use rB$_3$
Is Output
filled
No
Yes
Write One
Block on Tape
Initialize
Output Ctr
Incr. File 2
Use rB$_2$
2732
13
2446
12
Transfer
Is Output
filled
No
Yes
Write One
Block on Tape
Initialize
Output Ctr.
Incr. File 1
and File 2
Use rB$_1$, rB$_2$
2732
13
2732
B
Have we reached
The End of
Block File 2
Yes
2747
Is this the
last Cat$_1$ for this
Descriptor
Yes
CK for
E/T Sent
Yes
.7c
6v
4033
No
2547
Is this the
last cat$_1$ for this
Descriptor
No
4030
8v
No
Check for
E/T Sent.
Yes
4030
8v
No
.7b
6v
4033
Yes
.7c
6v.
4033
No
.7b
7v
4100

Fig. 15

Fig. 16


2745
14
Check for E/T Sent. File 2
No
.8b
.6e
2437
11-
Yes
16
2639
2625
15
2636
Is 8v .8b
No
.9b
.7c
4030
8v
Yes
16
2639
2639
16
Is the Output filled
Yes
2692
Write E/T Sent on Tape. File 2
2492
Rewind Servos for Files 2 & 3
.5c
.6a
.7a
.8a
.9a
Interchange Files 2 and 3
3v
2656
No
Fill last with P/B Sent. Write One Block on Tape

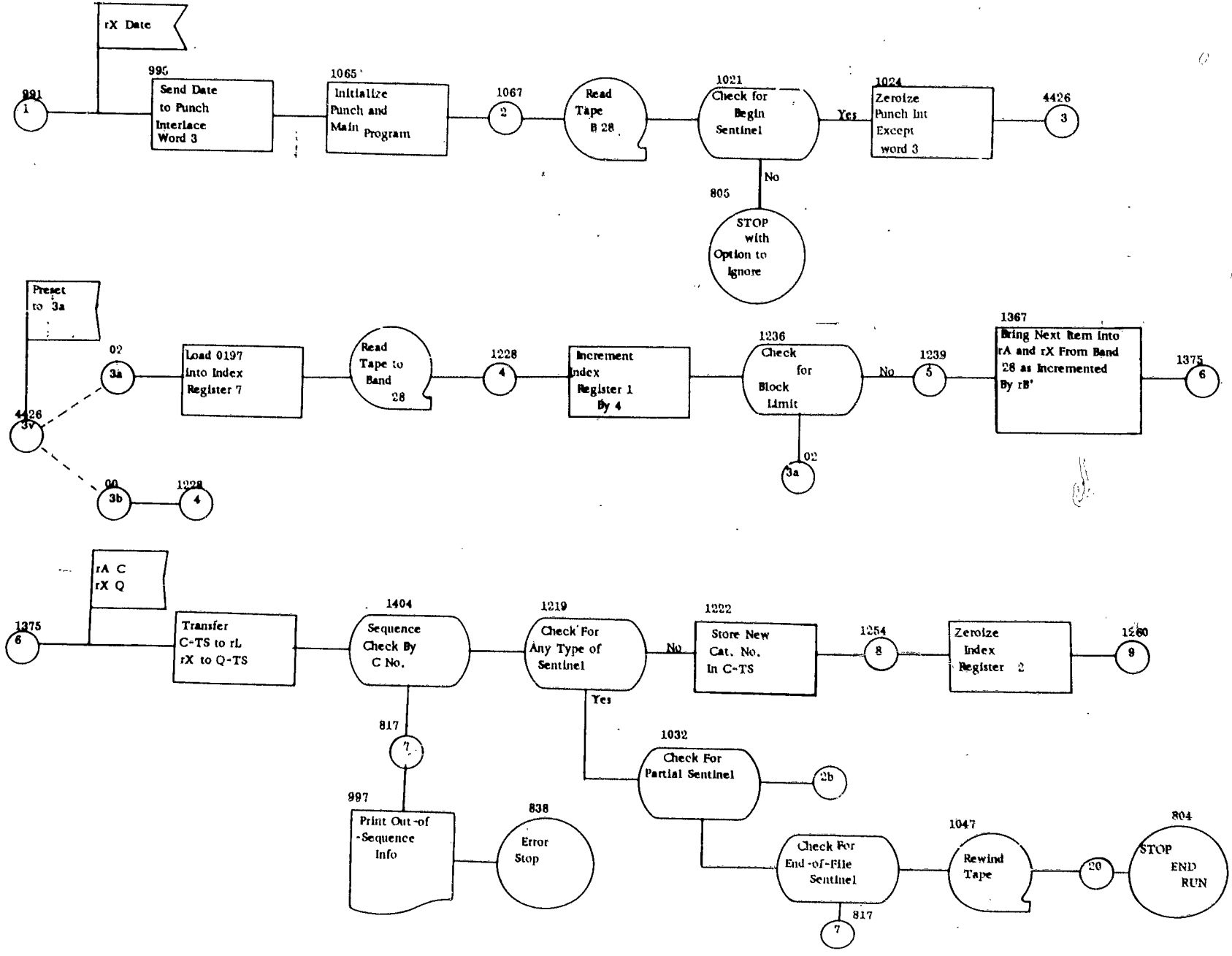

rX Date
991
1
995
Send Date to Punch Interlace Word 3
1065
Initialize Punch and Main Program
1067
2
Read Tape B 28
1021
Check for Begin Sentinel
Yes
1024
Zeroize Punch Int Except word 3
4426
3
No
805
STOP with Option to Ignore
Preset to 3a
02
3a
Load 0197 into Index Register 7
Read Tape to Band 28
1228
4
Increment Index Register 1 By 4
1236
Check for Block Limit
No
1239
5
1367
Bring Next Item into rA and rX From Band 28 as Incremented By rB'
1375
6
4426
3V
00
3b
1228
4
02
3a
rA C
rX Q
1375
6
Transfer C-TS to rL rX to Q-TS
1404
Sequence Check By C No.
1219
Check For Any Type of Sentinel
No
1222
Store New Cat. No. In C-TS
1254
8
Zeroize Index Register 2
1260
9
Yes
817
7
997
Print Out-of-Sequence Info
838
Error Stop
1032
Check For Partial Sentinel
2b
Check For End-of-File Sentinel
1047
Rewind Tape
20
804
STOP END RUN
817
7

Fig. 17

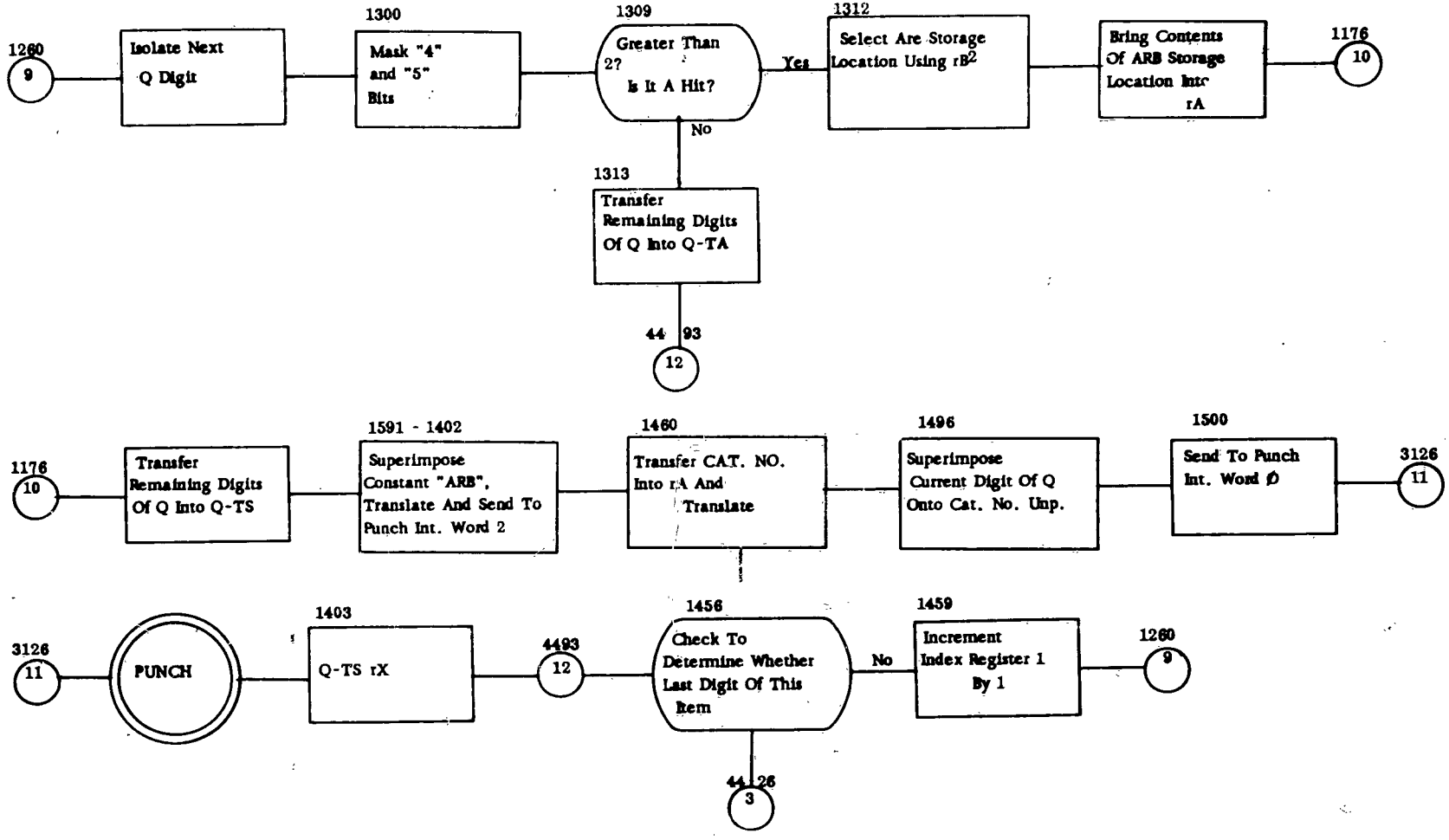

1260
9
Isolate Next Q Digit
1300
Mask "4" and "5" Bits
1309
Greater Than 2?
Is It A Hit?
Yes
No
1312
Select Are Storage Location Using rB2
Bring Contents Of ARB Storage Location Into rA
1176
10
1313
Transfer Remaining Digits Of Q Into Q-TA
44 93
12
1176
10
Transfer Remaining Digits Of Q Into Q-TS
1591 - 1402
Superimpose Constant "ARB", Translate And Send To Punch Int. Word 2
1460
Transfer CAT. NO. Into rA And Translate
1496
Superimpose Current Digit Of Q Onto Cat. No. Unp.
1500
Send To Punch Int. Word Ø
3126
11
3126
11
PUNCH
1403
Q-TS rX
4493
12
1456
Check To Determine Whether Last Digit Of This Item
No
1459
Increment Index Register 1 By 1
1260
9
44 26
3
786
Begin Recovery Routine
Zeroize
Band 28
C-TS
Q-TS
TS-Error RTN
630
Set 3a
Stop Recovery Complete

Fig. 18

RETRIEVAL PROGRAM MODIFICATIONS

The magnetic tape program for retrieval was designed initially to phase the production of mechanized bibliographies from the punch card into the tape operation. As can be expected in a dynamic situation, plans for expansion and refinement were underway before the program was ever employed operationally. Several program modifications for the tape operation are being introduced as the production schedule permits. These are in addition to considerations being given to employment of random access capacity.

The program was originally written to accommodate ten simultaneous searches with a maximum depth of four coordinations (identification of documents in the collection that contain a combination of four terms prescribed for a single search). In the actual machine operation, a group of terms can be treated as a single term. In certain types of combined searches, this greatly increases the efficiency of the machine operation. An illustration of this technique is given in the example on pages 32 and 33.

In running large searches employing a great number of terms, the records transferred to the output tapes increase to a point that it becomes more economical to remove the output tapes and continue the transfer of data on new output tapes. The partial

search tapes obtained in this manner are then merged to obtain the final output. In the merge, the search and coordination level identification data in the second word of the record are consolidated for matching AD numbers in the same manner as the merge on the output tapes of the primary run. In actual practice, it has proved to be profitable to break the output tape in this manner at about 400 blocks. Further study has indicated that for optimum operation, the merge now performed separately should be incorporated into the main run and programmed to transfer the data from output tapes automatically when the total blocks compiled exceed 100 blocks on completion of the transfer of a retrieval term record from the master tape.

In current practice, about eight hours is required to run ten average searches employing 120 - 130 unique retrieval terms. A 50% increase in efficiency can be expected in a search of this type by employment of the integrated compile-merge technique just described. In actual production, this will increase the capacity for searches from 10 to 15 for a single shift.

A second modification to the program will increase the capability from four to six levels of coordination. This modification will require only minor recoding of the output punch routine. This feature was not included in the program initially since it is seldom required to satisfy current customer

demands. When coordination of more than four terms are required, it is now necessary to make a second pass. Since this is seldom required, the four-level coordination is considered optimum for the current operation. It is believed that as customers become more willing to rely on the product of the machine retrieval system, a greater degree of specificity will be in demand. A sample input card for six-term coordinations is shown in Figure 19.

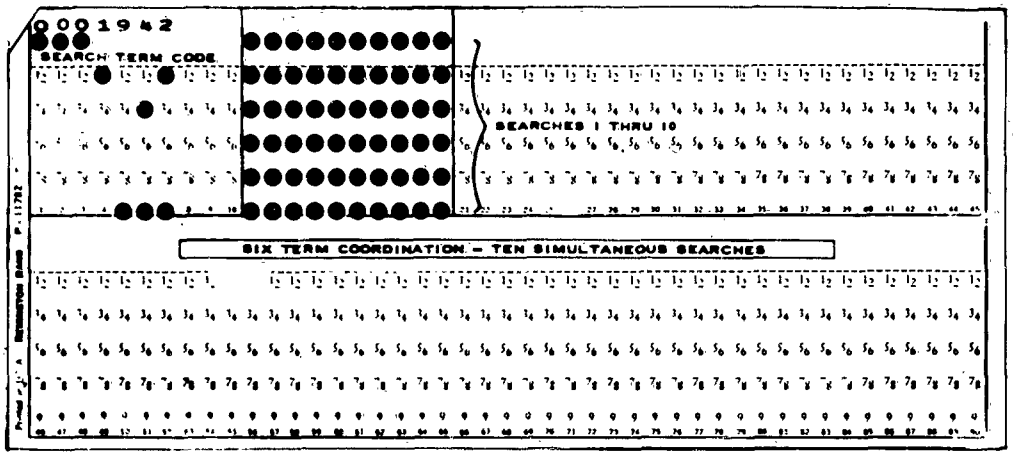


Fig. 19 SEARCH AND RETRIEVAL INPUT CARD

A third modification will increase the capacity for simultaneous searches. However, the level of coordination will be reduced as the number of searches increase beyond ten. A combined total of 60 retrieval term coordinations can be employed simultaneously. Any combination of coordinations can be employed as long as the total

does not exceed 60 (a group of terms can still be as a single term, however).

This type of capability will be extremely useful in compiling large state-of-the-art bibliographies where many two and three level searches are required to cover the bounds established by the reference analyst. In searches of this sort, many terms are common to more than one search. The greater the common occurrence, the greater the efficiency of the search since one pass of any retrieval term record will satisfy all searches that are combined in a single computer run. A sample input card shown in Figure 20 illustrates how twenty searches of three term coordinations each can be accommodated on a single input card.

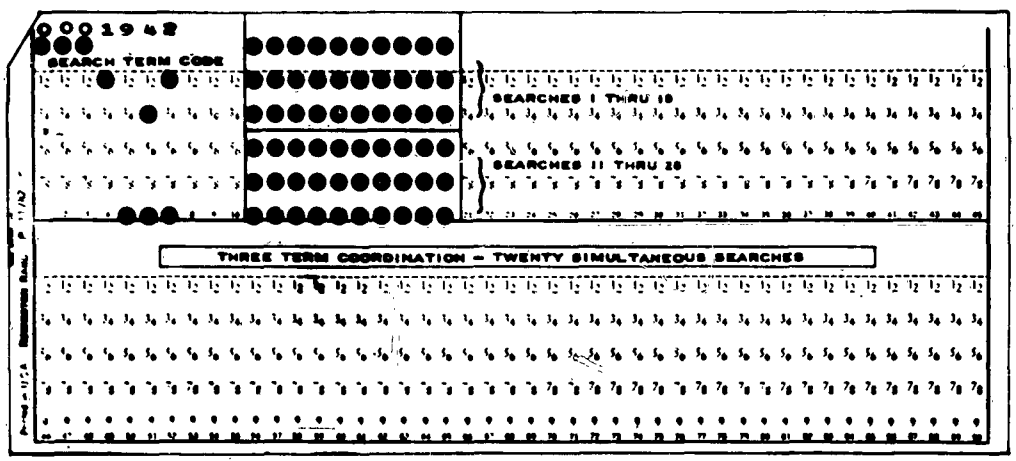


Fig. 20 SEARCH AND RETRIEVAL INPUT CARD

A fourth modification under consideration will not affect the search capacity of the run but should improve program efficiency, particularly on searches that employ a large number of retrieval terms (in excess of 50). In the present program, the two output work tapes must be rewound and reinitialized after merging of each record from the master tape. Although the computer is free for other operations during the rewinding, it is idle except for the time required to read the next input card and to reposition the master descriptor file tape. The amount of idle time will depend on the number and distribution of the term codes involved. In this modification, the output tapes will not rewind. The term records from the master tape will be merged on the output tapes alternating from low to high and high to low AD numbers. The first record from the master tape will be read forward from the beginning of the record to the end (low to high AD number); the second record will be read from the end back to the beginning (high to low AD number). This will eliminate all rewinding of the output tapes until the end of the search is reached. Although the idle time introduced by rewinding tapes will be greatly minimized under the integrated compile-merge techniques described earlier, it may still be more efficient to employ this low-high, high-low merge technique when several large retrieval term records are involved in the search.

There are many instances in the Thesaurus where two and three terms have been combined and assigned a single numeric code. A

review of the manual and machine experience over the past year indicates a great deal might be gained to increase this trend. Where the frequency of certain combinations of retrieval terms justifies assigning a single code to the combination, it will be done. The individual terms will still be retained in the system.

A separate program has been written to punch out the document listing under these combinations directly from the master tape file. An option can be exercised to punch out only the document listing required to update a previous run.

In the formative stages of the planning for automation, it was not considered economically feasible to employ random access for information retrieval. Since the early plans were formulated, however, several factors have prompted a new look at the random access operation. The Remington Rand UNIVAC RANDEX components have been transistorized, eliminating the separate power supply and consolidating the control unit into the tape synchronizer. The capacity has been expanded from 600,000 to 1,200,000 ten-digit words per RANDEX unit. This increased capacity is provided at slightly less rental.

At this stage of operation, it has not been determined how much, if any, computer time could be saved by placing the entire operation on RANDEX, including the program, master file, and search output. To what extent random access will be employed in search

and retrieval will depend on future developments and the optimum application of ASTIA's integrated automated capacity. One approach being considered at this time that shows considerable promise is a combination punch card--magnetic tape--random access operation. The input and output are on punch cards and the master retrieval file is on magnetic tape as in the present operation. The search and coordination level data from the input cards will be stored on the RANDEX drum at the address corresponding to the document numbers listed in the retrieval term records on the master retrieval file. The final search results stored on the RANDEX drum will be converted to punch cards. Initially, 400,000 words of RANDEX capacity will be utilized--one word to represent AD numbers 1 through 400,000. The AD number will be coincident with the RANDEX address and need not be stored on the RANDEX drum.

FUTURE CONSIDERATIONS IN SEARCH AND RETRIEVAL PROGRAMMING

Several possibilities for varying the weight of individual retrieval terms used in a search are under active consideration. It appeared to be desirable, from the start, to make some use of the descriptors selected as subject cumulators for compiling TAB indexes. In a sense, these descriptors are weighted because they describe the major subject area of a report. No use was made of the subject cumulators in the early magnetic tape programs because it was desired to keep the machine application as simple as possible. In the punch card production runs, coordination on descriptors that had been selected as subject cumulators was identified on the output punch cards as described earlier. No conclusive determination of the value of the subject cumulators to increase specificity of search had been made when this paper was written.

It has not yet been determined if a weighting system will actually be employed. The concept is to enable the bibliographic and reference personnel to use the same approach to retrieval as the Scientific Analysis personnel used when they assigned the retrieval terms to reports in the collection. The check list which serves as a guide for assigning retrieval terms to documents may or may not be useful as a guide to weighting individual search terms.

A further search refinement might be to identify a report by type, such as whether it is an initial report, progress report, final report, etc. It may even be feasible to weight the report on the basis of content. This latter consideration certainly is worthy of further investigation but may require a greater degree of coordination between ASTIA and the originators of reports added to the collection.

The only progress made to date on the concept of weighting descriptors has been to study the programming aspects. It was determined to be entirely feasible to provide for weighting retrieval terms in the current program logic. Only minor changes will have to be made in the tape run and in punching the output card from the tape.

In the master file, the AD number is carried in a one-word item thus: 0XXXXXX000. The use of a pattern of characters in the last three positions will produce discreet weighting codes. The first position in the word is reserved for control keys for manipulating and sorting the master file.

THE ROAD AHEAD

In the automatic data processing feasibility study ASTIA had suggested the use of punched paper tape equipment to prepare reproduction copy for the Technical Abstract Bulletin and catalog cards. However, the use of punched paper tape was not included in the original plan because at that time it simply represented an improvement in the method of preparing TAB and catalog card copy. It did not relate directly to data processing.

Shortly after the automation plan was approved and preparations had begun in earnest, it was decided to proceed with plans for introducing punched paper tape equipment into the copy preparation function. This decision was made because the EDP equipment which was finally selected had certain characteristics which could be capitalized on to advantage in terms of integrated data processing.

It became apparent that the punched paper tape which was to be created initially in the course of preparing reproducible copy could be converted to magnetic tape if proper provisions were made for such conversion. The information thus recorded on magnetic tape could be printed out on the high speed printer as a result of the bibliography search. In addition, much of the information which is to be captured in punched paper tape can be machine converted into

punched cards for the Index Master, Inventory Master and Information Retrieval files. In order to assure accuracy of the information, master edge punched cards are being used to enter certain authoritative information into the TAB copy and thus into paper tape.

The advantages of doing what has just been described are obvious. The duplicative tasks of key punching and verifying this same information could be eliminated. Here again it was necessary to face the fact that there is more systems design and actual programming to be done than the staff could handle all at once. Therefore, it has been necessary to phase into this aspect of integrated data processing on a time-available basis.

The program master tapes with special coding for the punch paper tape have been designed to permit the computer to vary format and select or omit data that have already been completed.

At this point, it is necessary to digress a bit and explain why program master tapes had to be developed for use with the punched paper tape equipment. The Technical Abstract Bulletin is an unclassified publication; yet it lists all documents cataloged, and some of these are classified as high as Secret. Consequently, it is essential that classified information be excluded from the TAB. On the other hand, bibliographies are geared to the requester's "need-to-know" and security clearance; hence, all the information concerning a particular document had to be captured in punched paper tape for subsequent use in bibliographies.

To accommodate both requirements, it is necessary to record all information on the punched paper tape. However, by using a program master tape, a by-product tape is produced which can be read by the machine at top speed to prepare the TAB-reproducible copy. The program master tape inserts control codes into the by-product tape in such a way as to suppress the classified information while machine typing the TAB reproducible copy. No by-product tape is punched in this latter process since the paper tape that is to be converted to magnetic tape is produced as a by-product tape at the time of typing the complete entry as described earlier.

In the early stages of punched paper tape planning, it was suggested by Mr. Laurence Buckland of ITEK Corporation that non-printing characters be employed as a programming aid to enable the computer to identify the various components of a complete document cataloging record. A suggestion along a similar line was also made by Mr. Calvin N. Mooers of the Zator Co. Mr. Mooers also advocated the employment of non-printing characters to separate the individual components of a record, but under a much broader concept. The objective under Mr. Mooers' concept was to provide a simple means for cooperation between libraries to improve cataloging operation and eliminate the necessity of multiple cataloging through a direct exchange of data in machinable form. If this objective can be realized, it will most certainly constitute a substantial advance in

the technology of documentation. A sample cataloging record compiled to the ASTIA format employing the non-printing characters compatible with Mr. Mooers' concept is shown on page 58. This record through program control can be printed from the punch paper tape in the format employed in the ASTIA Technical Abstract Bulletin. A sample is shown on page 59. Beginning with the 15 March 1961 issue, this procedure will be employed by ASTIA to compile the Technical Abstract Bulletin.

Implementation of the punched paper tape procedures began during November of 1960. To achieve a reasonable capability for machine printed bibliographies in a relatively short time, the cataloging information for documents now in the system is being selectively converted. At the same time all of the day-to-day input is being processed for TAB reproducible copy. In this way it is expected to attain a full capability for machine printing bibliographies within two years.

Another important factor in the use of the punched paper tape is that it will provide a medium for large scale experimenting with the computer in terms of information retrieval which goes beyond that based on descriptors assigned by the scientific analysts. There is much to be done in this field. A number of documentalists of national repute have already expressed hope that they will be able to utilize the potential which the punched paper tape project is creating.

ASTIA Data Identification System

▽ ▽ ▽ ▽ (1) AD-213 240L (2) ▽ ▽ Div. 6, 8, 9/2 (3) ▽ ▽ Unclassified (CR)
(4) ▽ ▽ (4 Nov 59) (CR)
(5) ▽ ▽ ↓ 083100 ↑ Arnold Engineering Development Center [Tullahoma, Tenn.] (CR)
(6) ▽ ▽ TEST OF AEROJET NOZZLES AND PROPELLANTS IN AN ALTITUDE TEST CELL FOR THE MINUTEMAN ICBM. PHASE IV, (CR) (CR)
(7) ▽ ▽ by Lawrence W. Kesting, F. E. Turner, Jr. and others, (8) ▽ ▽ Oct 59, (CR)
(9) ▽ ▽ 57p. incl. illus. tables (10) ▽ ▽ (Rept. no. AEDC TN 59-120) (CR)
(11) ▽ ▽ (Contract AF 40(600)800) (12) ▽ ▽ Unclassified report (CR)
(13) ▽ ▽ (AEND TN 4634) (CR)

(14) ▽ ▽ Notice: Only Military Offices may request from ASTIA. Others request approval of Arnold Engineering Development Center, Tullahoma, Tenn. (CR) (CR) (CR)

(15) ▽ ▽ Available on loan only (CR)

(16) ▽ ▽ DESCRIPTORS: * ↑ 114 271600 ↓ ↓ Guided missiles, ↑ * ↑ 173 5361000 ↓ ↓ Rocket motor nozzles, ↑ (CR)
↑ 027 606500 ↓ ↓ Surface to surface, ↑ ↑ 219 536000 ↓ ↓ Rocket propulsion, ↑ (CR)
↑ 085 329700 ↓ ↓ Jatos, ↑ ↑ 148 5799000 ↓ ↓ Specific impulse, ↑ ↑ 006 6302000 ↓ ↓ Thrust, ↑ (CR)
* ↑ 099 5740000 ↓ ↓ Solid rocket propellants, ↑ ↑ 149 028800 ↓ ↓ Altitude chambers. ↑ (CR)

(17) ▽ ▽ OPEN-ENDED TERMS: 123 1234567 L-24, 234 2345678 DP-104, 345 (CR)
3456789 VIP, 543 4567890 TR/C32. (18) ▽▽ (4 spaces) (CR)

(19) ▽ ▽ Ten Aerojet General Corporation modified 5KS 4500 JATO rocket motors were fired in an altitude test cell as part of a nozzle and propulsion development program in support of the Minuteman project. An average simulated pressure altitude above 106,000 it was maintained during the firings. The nozzle configurations were the "Demuth" contour and a 15 deg half angle conical with various throat curvatures and a 24:1 expansion ratio. Propellant formulation consisted of ammonium perchlorate, polyurethane, and aluminum. Static pressures were measured throughout the nozzle length. Average specific impulse ranged from 276 to 279 lb-sec/lb, based upon the manufacturer's statement of propellant weight, and from 261 to 274 lb-sec/lb, based upon the motor weight differences before and after firing. Nozzle pressures for the first 3 sec of burning agreed closely with theoretical values. (Author) (CR) (CR) (CR) (CR) (CR) (CR) (CR) (CR) (CR) (CR)
(20) ▽ ▽ (See also AD-200 000) ▽ ▽ ▽ ▽ ▽ (CR)

UNINTELLIGIBLE CODES

1-	▽	Formating Code	XS3 Code 237
2-	↑	Non Print Select End Code	XS3 Code 267
3-	↓	Print Restore Code	XS3 Code 34567
4-		Shift Code	XS3 Code 457
5-		Shift Release Code	XS3 Code 12457
5-	(CR)	Carriage Return Code	XS3 Code 3

AD-108 290 Div. 13

Armour Research Foundation, Chicago, Ill.
EFFECT OF LONG VERSUS SHORT DURATION BLAST LOADINGS ON STRUCTURES.
by Alfred Ritter and T. H. Schiffman, 22 Aug 55, 21p. incl. illus. tables (Rept. no. 18; Supersedes rept. no. 18, AD-72 567) Final test rept. no. 1 on Blast Effects on Buildings and Structures: Operation of Six-Foot and Two-foot Shock tubes.
(Contract AF 33(616)2644) Unclassified report

Descriptors: Structures, Blast, Vulnerability, Pressure, Tests, Deflection,

''An exploratory investigation was conducted to demonstrate on a gross basis the reduction in peak overpressure which will produce a given maximum deflection of an elastic-plastic structure as the positive phase duration of blast loading is increased. Test structures were subjected to blast waves of approximately 10 and 20 msec positive phase durations and approximately 20 and 10 psi peak overpressures, respectively. Although no quantitative conclusions could be drawn as to exact pressure reduction factors because of limited accuracy and small number of tests, the results indicate substantial pressure deductions with increased duration loadings for the test structures under consideration.'' (Contractor's abstract) (See also AD-51 982)

AD-153 545 Div. 13, 22, 30

Ballistic Research Labs., Aberdeen Proving Ground, Md.
AIR BLAST LOADING ON THREE-DIMENSIONAL SCALE MODELS OF A SEMI-CYLINDER,
by F. B. Smith, Jr., E. G. Rines, and J. H. Keefer. July 57, 84p. incl. illus. tables (Proj. no. TB3-0112; BRL memo rept. no. 1092: AFSWP no. 1047)
Unclassified report

Descriptors: Air blast, Recording devices, Blast, Shock waves, Diffraction, Structures,

The objective of this investigation was to record the diffraction loading on a scaled model exposed in the shock tube. The model used was a 1/72 scale of the Navy Structure 3.2.5 exposed on Operation Greenhouse. The work was requested by the Bureau of Yards and Docks, Department of the Navy, and carried out as an extension of AFSWP Project 3.28.1 of Upshot/Knothole. Pressure vs time curves of the diffraction phase of the air blast loading at 28 positions, 3 orientations on 3 shock strengths are presented in this report. The curves are presented so that comparisons may be made with results of full scale tests and others shock tube experiments. (Author) (See also AD-74 974)

The results are present
data acquisition method
of high-altitude nucles
telemetry, and recovery
tained through literatu
data searches, visits t
and independent researc
study surveys the past
methods, describes rece
suggests new systems. A
losophy which evolved f
sented, along with the
lection and design whic

AD-238 039 Div. 22,
(20 June 60)

Ballistic Research Labs
Ground, Md.
THE RELATION OF SEISMIC
THE AREA UNDER THE STRE
by A. A. Thompson, Apr
(DA Proj. 5B03-04-002
BRL memo. rept. no. 126

Descriptors: Mathematic
*Explosions, Attenuati
*Seismic waves,

Seismic energy attenuat
was determined by measu
various distances from
also by measuring the a
and unloading parts of
Theory and experiment s
attenuation determined
curve may be less than
attenuation is very lar

AD-239 871 Div. 13,
(19 July 60)

Naval Civil Engineering
Calif.
LITERATURE SURVEY OF CO
RADIATION SHIELDING,
by Meldon Merrill and W
31p. incl. illus. 41 re
Type C (Proj. Y-RO 11-0
no. 084)

Descriptors: Magnetite,
Bibliography, Sulfates
Barium compounds, Mine
Additives, *Reactor sh
*Concrete,

CONCLUSIONS: The use of
increasing the density
providing a more effici

Sample Page Format
Technical Abstract Bulletin

SUMMARY OF PRODUCTION AND PROGRAMMING TASKS

The major phases of production and programming effort involved in the creation of an automated information search and retrieval system have been covered in this publication and in the two previous publications on ASTIA's automation program, AD-227 000 and AD-247 000. It is believed that a brief summary of the various tasks involved in establishing the operational system is in order. This summary lists those tasks that involve a significant amount of production effort. In some instances, however, creativity and ingenuity were more significant than the expenditure of production resources.

Production Task	Remark
1. Compile thesaurus of descriptors.	This task was completed by ASTIA personnel and published in May 1960. It is described in AD-227 000.
2. Develop numeric coding system for descriptors.	This task was completed in August 1959.
3. Assign descriptors to documents in the AD collection.	Approximately 205,000 reports were in the AD collection when the project became current. The collection is now in excess of 240,000.
4. Code descriptors in accordance with coding system developed under Production Task 2 above.	Accomplished concurrently with Production Task 3. Became current 1 June 1960.
5. Key punch retrieval term cards for reports in the AD collection.	Key punching for reports cataloged prior to 1 May 1960 was accomplished by contractor. (1,200,000 cards) Subsequent to 1 May 1960, key punching accomplished in-house. Current workload: 1,000 cards a day.

Production Task	Remark
6. Prepare publications and instructions on ASTIA automation system.	AD-227 000, AD-247 000, and this publication.
7. Sort retrieval term cards to separate indexing terms and open-ended terms.	Maintained on current basis.
8. Alphabetic sort of open-ended terms.	Required initially for standardization and clean-up. Maintained on current basis for new input.
9. Print tabulation of open-ended terms.	Required initially for standardization and clean-up. Project completed.
10. Reproduce and sort retrieval term cards for testing system and production of bibliographies on card computer.	From AD-150,000 to reports announced through 15 July 1960. Later expanded to include reports announced through 15 November 1960. File is still maintained for limited use. It consists of 300,000 cards.
11. Convert retrieval term cards to magnetic tape in 12-word item.	In-house project accomplished during system test period. 12-word item used to take advantage of available program. (1,400,000 cards involved.)
12. Reformat tape retrieval term file (product of Production Task 11) to 100-word item. (50 two-word records to a block.)	This reformat also placed AD number and descriptor code in proper position within the record for two-word sort.
13. Sort retrieval term tapes (product of Production Task 12) into descriptor code primary and AD number secondary.	Step in preparation of master descriptor tape file for retrieval application.
14. Sequence check sorted retrieval term tapes. (Product of Production Task 13.)	Required for clean-up and to remove non-sortable data from file and to write data not in sequence on work tape to be resorted.

Production Task	Remark
15. Merge sorted retrieval term tapes (product of Production Task 14) into continuous file.	Five 2400-foot tapes involved. Production Tasks 14 and 15 were later consolidated and accomplished in a combined operation.
16. Reformat merged retrieval term tapes from two-word record, 100-word item, into 50-word item.	This is format of master retrieval term file for operational runs.
17. Sequence check master retrieval term tape file. (Product of Production Task 16.)	Required to maintain sequence integrity of file. This task must also be performed after each twice-monthly updating of file.
18. Print out final master retrieval term tape file.	Required for data clean-up and visual reference.
19. Key punch numeric codes for open-ended terms.	Step in integrating open-ended terms into system.
20. Convert open-ended terms from cards to magnetic tape. Sort to descriptor code sequence primary and AD number secondary.	Step in integrating open-ended terms into system. This will be accomplished when more experience has been gained in employment of open-ended terms. The card file is now employed in original format.
21. Sort indexing terms (descriptors used as subject cumulators) of entire AD collection into descriptor code primary and AD secondary. Merge identifying code into master descriptor file.	The addition of these data to the file will provide a capability to index large bibliographies, or possibly the entire AD collection. It is also possible that they may be employed as weighting factors in the actual retrieval runs.
22. Update master retrieval term tape file twice-monthly. (Follow by sequence check, Production Task 17.)	This is a card-to-tape merge. Cards are sorted off-line to the tape sequence. Approximately 11,000 cards are involved in each updating. (Deletions may also be made in same run.)

Production Task	Remark
23. Preparation of test data for computer programs involved in establishing retrieval system.	Programs are described individually.
24. Operational testing.	Over 400 bibliographies were run on a production basis and checked manually.
25. Reformat two-word IR file (product of Production Task 12) from descriptor-AD to AD-descriptor.	Required to use existing two-word sort to establish file in AD sequence. Reformat run also block sorts on AD number in same operation.
26. Sort IR file (product of Production Task 25) into AD primary, descriptor secondary.	Required for data clean-up of IR file.
27. Print out gaps in AD listing in IR file.	Required for data clean-up of IR file.
28. Print out gaps in AD listing in inventory file.	Task already accomplished, required for manual check against product of Production Task 27 for data clean-up of IR file.
29. Statistical run of IR file to determine frequency of use of descriptors.	Statistical data required to optimize operational runs. Product is in two-word record thus: 00000XXXXX (Number of AD's) 000XXXXXXX (Descriptor)
30. Sort product of Production Task 29.	Two-word sort employed to convert IR file to descriptor frequency sequence. Required to optimize operational runs.
31. Reconstitute IR file in descriptor frequency sequence.	Required to optimize operational runs.
32. Compile descriptor combination frequency of operational runs.	Required to provide data for consolidation records of frequently combined records to improve efficiency of system.

PROGRAMMING

Most of these programs were started before "canned" routines were available. The entrance and exit memory locations were furnished by the manufacturer. Many of the programs have been modified or rewritten to provide more efficient operation for particular applications.

Programming Tasks	Remark
1. Conversion of retrieval term cards to magnetic tape in 12-word item.	This format employed to take advantage of tape control routine then available.
2. Reformat retrieval term tape file from 12-word item to 100-word item. (50 two-word records to a block of tape.)	Needed for format tape for sorting.
3. Sort two-word record retrieval term file.	Modification of 12-word sort prepared by DATATROL Corporation under contract to Remington Rand UNIVAC at no cost to ASTIA.
4. Sequence check two-word record retrieval tape file.	Required for data clean-up prior to creation of operational file.
5. Merge two-word record retrieval term file into one continuous record.	Later modified to accomplish clean-up incorporated in Program Task 3.
6. Reformat sorted and merged two-word retrieval file into two 50-word items.	Required to create operational master retrieval term tape file.
7. Print out two-word record retrieval term file.	Needed to test programs and verify data during several production operations in creating file.
8. Print out 50-word item.	Needed to print out master retrieval file and to test programs and verify data.

Programming Tasks	Remark
9. Convert card-to-tape in 50-word item to simulate master file.	Needed to debug and test operational retrieval run, Part I.
10. Convert card-to-tape two-word item to simulate output from retrieval run, Part I.	Needed to debug and test operational retrieval run, Part II.
11. Merge retrieval term cards into 50-word item master retrieval term tapes.	Needed for twice-monthly updating of master retrieval term file.
12. Operational retrieval run -- 6-level coordination.	Initial programming approach. Consists of five runs and three sorts.
13. Operational retrieval run -- 4-level, ten simultaneous searches.	Part I: Tape-to-tape run. Part II: Tape-to-punched cards run.
14. Operational retrieval run -- 6-level, ten simultaneous searches.	Modification of Programming Task 13, Part II.
15. Operational retrieval run -- 60-descriptor (or descriptor group coordination).	Modification of Programming Task 13, Part II.
16. Operational retrieval run -- high low-low high tape merge.	Modification of Programming Task 13, Part I. Speeds up computer run.
17. Merge-buff, two-word.	Used to combine sub-searches -- increases program efficiency.
18. Combine two-word merge buff into retrieval run, Part I.	Integrates product of Programming Task 17 into retrieval run, Part I, to optimize output.
19. Reformat and block sort two-word item IR tape.	Needed to reformat IR file for two-word sort to AD sequence primary.
20. Print out gaps in AD listing of IR file.	Required for data clean-up of IR file.

Programming Tasks	Remark
21. Punch out single descriptor records for 50-word item IR tape.	Required for production runs where descriptor combinations or single descriptor records are required.
22. Modify Programming Task 21 to punch out only partial references to update previous punch-out.	
23. Count AD listing in descriptor record and compile in two-word record. 00000XXXXX 000XXXXXXX Number of AD's Descriptor	To provide data of distribution of descriptors.